BI-VZD přednáška 6

Alexander Kovalenko

FIT ČVUT

28. 03. 2022

Autoři: Karel Klouda, Juan Pablo Maldonado Lopez, Daniel Vašata.
Problémy, návrhy apod. hlaste v GitLabu.
Verze souboru: 28. března 2022 10:41.

Co bude v dnešní přednášce

- Připomenutí lineární regrese
- Geometrická interpretace metody nejmenších čtverců
- Problém lineárně závislých sloupců
- Regularizace pomocí hřebenové regrese
- Modely bázových funkcí

Lineární model

- Model pro vysvětlovanou proměnnou Y v bodě $\boldsymbol{x}$ je

$$Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon = w_0 + w_1x_1 + \ldots w_px_p + \varepsilon.$$

Lineární model

- Model pro vysvětlovanou proměnnou Y v bodě $\boldsymbol{x}$ je

$$Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon = w_0 + w_1x_1 + \ldots w_px_p + \varepsilon.$$

- Model pro trénovací množinu tvořenou N páry $(Y_i, \boldsymbol{x}_i)$ je $Y_i = \boldsymbol{x}_i^T\boldsymbol{w} + \varepsilon_i$.

Lineární model

- Model pro vysvětlovanou proměnnou Y v bodě $\boldsymbol{x}$ je

$$Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon = w_0 + w_1x_1 + \dots w_px_p + \varepsilon.$$

- Model pro trénovací množinu tvořenou N páry $(Y_i, \boldsymbol{x}_i)$ je $Y_i = \boldsymbol{x}_i^T\boldsymbol{w} + \varepsilon_i$.
- Dohromady při značení $\boldsymbol{x}_i = (1, x_{i;1}, \dots, x_{i;p})^T$ tedy

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & x_{1;2} & \cdots & x_{1;p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & x_{N;2} & \cdots & x_{N;p} \end{pmatrix} \begin{pmatrix} w_0 \\ \vdots \\ w_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Lineární model

- Model pro vysvětlovanou proměnnou Y v bodě $\boldsymbol{x}$ je

$$Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon = w_0 + w_1x_1 + \dots w_px_p + \varepsilon.$$

- Model pro trénovací množinu tvořenou N páry $(Y_i, \boldsymbol{x}_i)$ je $Y_i = \boldsymbol{x}_i^T\boldsymbol{w} + \varepsilon_i$.
- Dohromady při značení $\boldsymbol{x}_i = (1, x_{i;1}, \dots, x_{i;p})^T$ tedy

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & x_{1;2} & \cdots & x_{1;p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & x_{N;2} & \cdots & x_{N;p} \end{pmatrix} \begin{pmatrix} w_0 \\ \vdots \\ w_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Toto zapisujeme maticově jako $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & \cdots & x_{1;p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & \cdots & x_{N;p} \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Lineární model

- Model pro vysvětlovanou proměnnou Y v bodě $\boldsymbol{x}$ je

$$Y = \boldsymbol{w}^T\boldsymbol{x} + \varepsilon = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon = w_0 + w_1x_1 + \ldots w_px_p + \varepsilon.$$

- Model pro trénovací množinu tvořenou N páry $(Y_i, \boldsymbol{x}_i)$ je $Y_i = \boldsymbol{x}_i^T\boldsymbol{w} + \varepsilon_i$.
- Dohromady při značení $\boldsymbol{x}_i = (1, x_{i;1}, \ldots, x_{i;p})^T$ tedy

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & x_{1;2} & \cdots & x_{1;p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & x_{N;2} & \cdots & x_{N;p} \end{pmatrix} \begin{pmatrix} w_0 \\ \vdots \\ w_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Toto zapisujeme maticově jako $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{1;1} & \cdots & x_{1;p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N;1} & \cdots & x_{N;p} \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Naměřené hodnoty jednotlivých příznaků $X_1, \ldots, X_p$ spolu s přidaným umělým příznakem $X_0 = 1$ jsou tedy uvedeny ve sloupcích matice $\mathbf{X}$.

Metoda nejmenších čtverců

- Při trénování minimalizujeme residuální součet čtverců

$$\mathrm{RSS}(\boldsymbol{w}) = \sum_{i=1}^{N} (Y_i - \boldsymbol{x}_i^T \boldsymbol{w})^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2.$$

Metoda nejmenších čtverců

- Při trénování minimalizujeme residuální součet čtverců

$$\mathrm{RSS}(\boldsymbol{w}) = \sum_{i=1}^{N} (Y_i - \boldsymbol{x}_i^T \boldsymbol{w})^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2.$$

- Minimum je určeno řešením **normální rovnice**

$$\mathbf{X}^T \boldsymbol{Y} - \mathbf{X}^T \mathbf{X} \boldsymbol{w} = \mathbf{0},$$

  která odpovídá podmínce $\nabla \mathrm{RSS}(\boldsymbol{w}) = \mathbf{0}$.

Metoda nejmenších čtverců

- Při trénování minimalizujeme residuální součet čtverců

$$\mathrm{RSS}(\boldsymbol{w}) = \sum_{i=1}^{N} (Y_i - \boldsymbol{x}_i^T \boldsymbol{w})^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2.$$

- Minimum je určeno řešením **normální rovnice**

$$\mathbf{X}^T \boldsymbol{Y} - \mathbf{X}^T \mathbf{X} \boldsymbol{w} = \mathbf{0},$$

  která odpovídá podmínce $\nabla \mathrm{RSS}(\boldsymbol{w}) = \mathbf{0}$.

- Za předpokladu, že je matice $\mathbf{X}^T\mathbf{X}$ regulární, existuje jediné řešení minimalizující $\mathrm{RSS}(\boldsymbol{w})$,

$$\hat{\boldsymbol{w}}_{\mathrm{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{Y}.$$

Metoda nejmenších čtverců

- Při trénování minimalizujeme residuální součet čtverců

$$\mathrm{RSS}(\boldsymbol{w}) = \sum_{i=1}^{N} (Y_i - \boldsymbol{x}_i^T \boldsymbol{w})^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2.$$

- Minimum je určeno řešením **normální rovnice**

$$\mathbf{X}^T \boldsymbol{Y} - \mathbf{X}^T \mathbf{X} \boldsymbol{w} = \mathbf{0},$$

  která odpovídá podmínce $\nabla \mathrm{RSS}(\boldsymbol{w}) = \mathbf{0}$.

- Za předpokladu, že je matice $\mathbf{X}^T\mathbf{X}$ regulární, existuje jediné řešení minimalizující $\mathrm{RSS}(\boldsymbol{w})$,

$$\hat{\boldsymbol{w}}_{\mathrm{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{Y}.$$

- Predikce v bodě $\boldsymbol{x}$ je potom $\hat{Y} = \boldsymbol{x}^T \hat{\boldsymbol{w}}_{\mathrm{OLS}}$.

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.
- To znamená, že pro optimální $\boldsymbol{w}$ je Eukleidovská vzdálenost bodů $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$ v prostoru $\mathbb{R}^N$ nejmenší možná.

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.
- To znamená, že pro optimální $\boldsymbol{w}$ je Eukleidovská vzdálenost bodů $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$ v prostoru $\mathbb{R}^N$ nejmenší možná.
- Označíme-li i-tý sloupec matice $\mathbf{X}$ jako $\mathbf{X}_{\bullet i}$, můžeme si všimnout, že

$$\mathbf{X}\boldsymbol{w} = w_0\mathbf{X}_{\bullet 0} + w_1\mathbf{X}_{\bullet 1} + \ldots + w_p\mathbf{X}_{\bullet p}.$$

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.
- To znamená, že pro optimální $\boldsymbol{w}$ je Eukleidovská vzdálenost bodů $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$ v prostoru $\mathbb{R}^N$ nejmenší možná.
- Označíme-li i-tý sloupec matice $\mathbf{X}$ jako $\mathbf{X}_{\bullet i}$, můžeme si všimnout, že

$$\mathbf{X}\boldsymbol{w} = w_0\mathbf{X}_{\bullet 0} + w_1\mathbf{X}_{\bullet 1} + \ldots + w_p\mathbf{X}_{\bullet p}.$$

- Vektor $\mathbf{X}\boldsymbol{w}$ je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty $w_0, \ldots, w_p$.

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.
- To znamená, že pro optimální $\boldsymbol{w}$ je Eukleidovská vzdálenost bodů $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$ v prostoru $\mathbb{R}^N$ nejmenší možná.
- Označíme-li i-tý sloupec matice $\mathbf{X}$ jako $\mathbf{X}_{\bullet i}$, můžeme si všimnout, že

$$\mathbf{X}\boldsymbol{w} = w_0\mathbf{X}_{\bullet 0} + w_1\mathbf{X}_{\bullet 1} + \ldots + w_p\mathbf{X}_{\bullet p}.$$

- Vektor $\mathbf{X}\boldsymbol{w}$ je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty $w_0, \ldots, w_p$.
- Leží tedy v lineárním podprostoru prostoru $\mathbb{R}^N$, který je lineárním obalem $p+1$ sloupců $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$.

Geometrická interpretace metody nejmenších čtverců (1/3)

- Minimalizace $\mathrm{RSS}(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2$ je ekvivalentní minimalizaci $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$.
- To znamená, že pro optimální $\boldsymbol{w}$ je Eukleidovská vzdálenost bodů $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$ v prostoru $\mathbb{R}^N$ nejmenší možná.
- Označíme-li i-tý sloupec matice $\mathbf{X}$ jako $\mathbf{X}_{\bullet i}$, můžeme si všimnout, že

$$\mathbf{X}\boldsymbol{w} = w_0\mathbf{X}_{\bullet 0} + w_1\mathbf{X}_{\bullet 1} + \ldots + w_p\mathbf{X}_{\bullet p}.$$

- Vektor $\mathbf{X}\boldsymbol{w}$ je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty $w_0, \ldots, w_p$.
- Leží tedy v lineárním podprostoru prostoru $\mathbb{R}^N$, který je lineárním obalem $p+1$ sloupců $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$.
- Pro různé hodnoty $\boldsymbol{w}$ pak vektor $\mathbf{X}\boldsymbol{w}$ celý tento prostor pokrývá, tj.

$$\langle \mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p} \rangle = \{\mathbf{X}\boldsymbol{w} \mid \boldsymbol{w} \in \mathbb{R}^{p+1}\}.$$

Geometrická interpretace metody nejmenších čtverců (2/3)

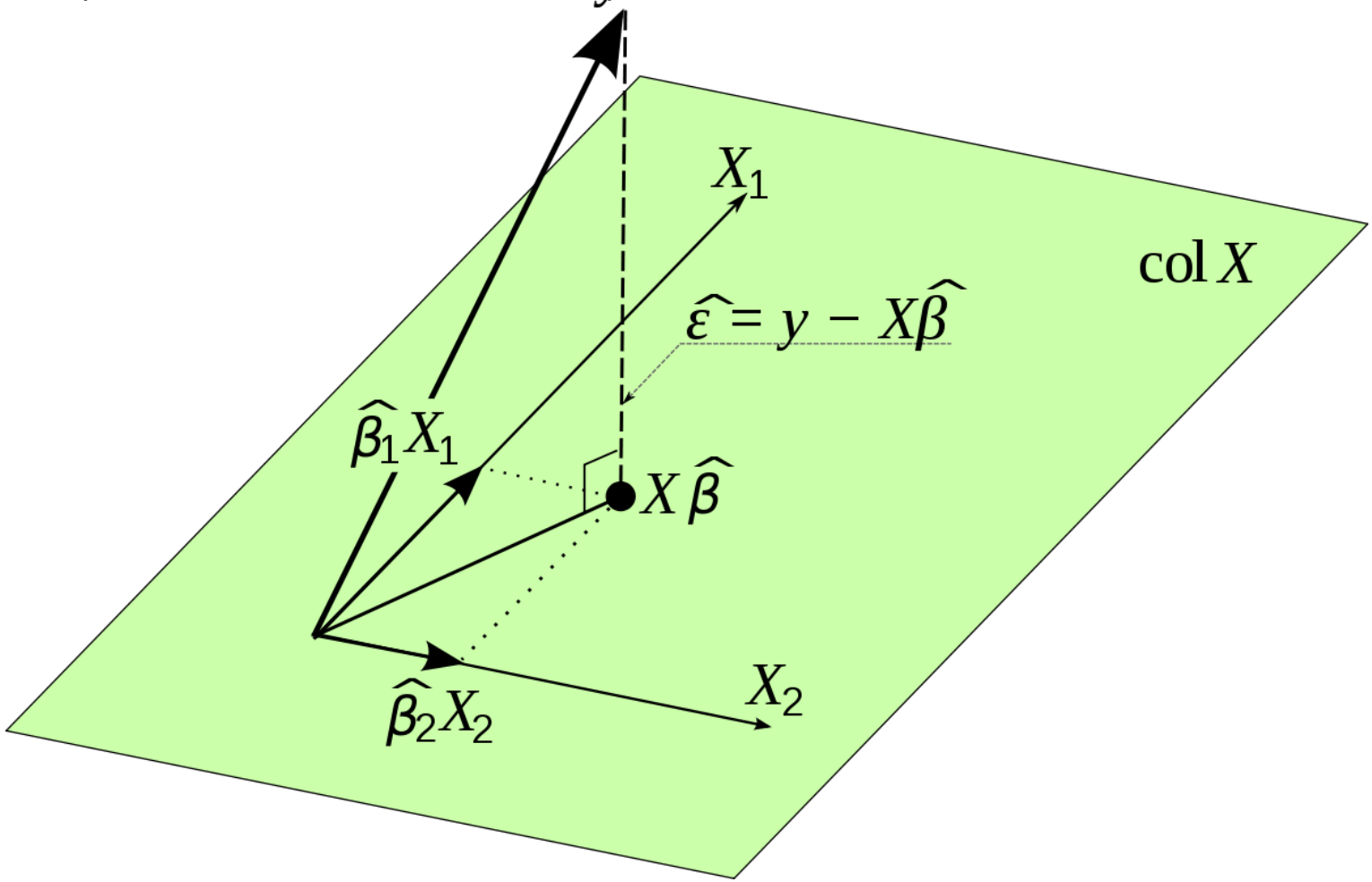
- Chceme-li minimalizovat vzdálenost $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$, hledáme bod $\mathbf{X}\boldsymbol{w}$ v podprostoru sloupců matice $\mathbf{X}$, který je k $\boldsymbol{Y}$ nejblíže.

Geometrická interpretace metody nejmenších čtverců (2/3)

- Chceme-li minimalizovat vzdálenost $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$, hledáme bod $\mathbf{X}\boldsymbol{w}$ v podprostoru sloupců matice $\mathbf{X}$, který je k $\boldsymbol{Y}$ nejblíže.
- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý (modrý vektor na obrázku).

Geometrická interpretace metody nejmenších čtverců (2/3)

- Chceme-li minimalizovat vzdálenost $\boldsymbol{Y}$ a $\mathbf{X}\boldsymbol{w}$, hledáme bod $\mathbf{X}\boldsymbol{w}$ v podprostoru sloupců matice $\mathbf{X}$, který je k $\boldsymbol{Y}$ nejblíže.
- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý (modrý vektor na obrázku).



Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.

Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.
- To znamená, že je kolmý na všechny vektory $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$, které ho generují:

$$(\mathbf{X}_{\bullet i})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = 0 \quad \text{pro všechny} \quad i = 0, \ldots, p.$$

Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.
- To znamená, že je kolmý na všechny vektory $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$, které ho generují:

$$(\mathbf{X}_{\bullet i})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = 0 \quad \text{pro všechny} \quad i = 0, \ldots, p.$$

- To lze maticově zapsat jako

$$\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = \mathbf{0} \quad \text{a tedy} \quad \mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}.$$

Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.
- To znamená, že je kolmý na všechny vektory $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$, které ho generují:

$$(\mathbf{X}_{\bullet i})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = 0 \quad \text{pro všechny} \quad i = 0, \ldots, p.$$

- To lze maticově zapsat jako

$$\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = \mathbf{0} \quad \text{a tedy} \quad \mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}.$$

- Získali jsme tím starou známou normální rovnici a tedy stejné řešení.

Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.
- To znamená, že je kolmý na všechny vektory $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$, které ho generují:

$$(\mathbf{X}_{\bullet i})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = 0 \quad \text{pro všechny} \quad i = 0, \ldots, p.$$

- To lze maticově zapsat jako

$$\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = \mathbf{0} \quad \text{a tedy} \quad \mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}.$$

- Získali jsme tím starou známou normální rovnici a tedy stejné řešení.
- Z výše uvedených geometrických úvah navíc plyne, že pro jakékoliv řešení $\boldsymbol{w}$ normální rovnice je $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$ a tedy i $\mathrm{RSS}(\boldsymbol{w})$ nejmenší možné.

Geometrická interpretace metody nejmenších čtverců (3/3)

- Bod $\mathbf{X}\boldsymbol{w}$ je k bodu $\boldsymbol{Y}$ nejblíže, jestliže je vektor $\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}$ na ten podprostor kolmý.
- To znamená, že je kolmý na všechny vektory $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$, které ho generují:

$$(\mathbf{X}_{\bullet i})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = 0 \quad \text{pro všechny} \quad i = 0, \ldots, p.$$

- To lze maticově zapsat jako

$$\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) = \mathbf{0} \quad \text{a tedy} \quad \mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}.$$

- Získali jsme tím starou známou normální rovnici a tedy stejné řešení.
- Z výše uvedených geometrických úvah navíc plyne, že pro jakékoliv řešení $\boldsymbol{w}$ normální rovnice je $\|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|$ a tedy i $\mathrm{RSS}(\boldsymbol{w})$ nejmenší možné.
- Jakékoliv řešení normální rovnice tedy dává globální minimum.

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.
- Pojďme si odvodit, jak to souvisí s lineární nezávislostí sloupců matice $\mathbf{X}$.

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.
- Pojďme si odvodit, jak to souvisí s lineární nezávislostí sloupců matice $\mathbf{X}$.
- Je-li $\mathbf{X}_{\bullet i}$ i-tý sloupec matice $\mathbf{X}$, platí, že vektor

$$\mathbf{X}\boldsymbol{s} = s_0\mathbf{X}_{\bullet 0} + s_1\mathbf{X}_{\bullet 1} + \ldots + s_p\mathbf{X}_{\bullet p}$$

je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty danými složkami $\boldsymbol{s}$.

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.
- Pojďme si odvodit, jak to souvisí s lineární nezávislostí sloupců matice $\mathbf{X}$.
- Je-li $\mathbf{X}_{\bullet i}$ i-tý sloupec matice $\mathbf{X}$, platí, že vektor

$$\mathbf{X}\boldsymbol{s} = s_0\mathbf{X}_{\bullet 0} + s_1\mathbf{X}_{\bullet 1} + \ldots + s_p\mathbf{X}_{\bullet p}$$

  je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty danými složkami $\boldsymbol{s}$.

- Matice $\mathbf{X}$ má lineárně nezávislé sloupce, právě když je $\mathbf{X}\boldsymbol{s} = \mathbf{0}$ pouze pro $\boldsymbol{s} = \mathbf{0}$.

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.
- Pojďme si odvodit, jak to souvisí s lineární nezávislostí sloupců matice $\mathbf{X}$.
- Je-li $\mathbf{X}_{\bullet i}$ i-tý sloupec matice $\mathbf{X}$, platí, že vektor

$$\mathbf{X}\boldsymbol{s} = s_0\mathbf{X}_{\bullet 0} + s_1\mathbf{X}_{\bullet 1} + \ldots + s_p\mathbf{X}_{\bullet p}$$

  je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty danými složkami $\boldsymbol{s}$.

- Matice $\mathbf{X}$ má lineárně nezávislé sloupce, právě když je $\mathbf{X}\boldsymbol{s} = \mathbf{0}$ pouze pro $\boldsymbol{s} = \mathbf{0}$.

- Obecně pro matici $\mathbf{X}$ a libovolný vektor $\boldsymbol{s} \in \mathbb{R}^{p+1}$ platí

$$\mathbf{X}\boldsymbol{s} = \mathbf{0} \;\Rightarrow\; \mathbf{X}^T\mathbf{X}\boldsymbol{s} = \mathbf{0} \;\Rightarrow\; \boldsymbol{s}^T\mathbf{X}^T\mathbf{X}\boldsymbol{s} = 0 \;\Rightarrow\; \|\mathbf{X}\boldsymbol{s}\|^2 = 0 \;\Rightarrow\; \mathbf{X}\boldsymbol{s} = \mathbf{0}.$$

Regularita versus lineární nezávislost sloupců (1/3)

- Normální rovnice má jednoznačné řešení, pokud je $\mathbf{X}^T\mathbf{X}$ regulární.
- Pojďme si odvodit, jak to souvisí s lineární nezávislostí sloupců matice $\mathbf{X}$.
- Je-li $\mathbf{X}_{\bullet i}$ i-tý sloupec matice $\mathbf{X}$, platí, že vektor

$$\mathbf{X}\boldsymbol{s} = s_0\mathbf{X}_{\bullet 0} + s_1\mathbf{X}_{\bullet 1} + \ldots + s_p\mathbf{X}_{\bullet p}$$

  je lineární kombinací sloupců matice $\mathbf{X}$ s koeficienty danými složkami $\boldsymbol{s}$.

- Matice $\mathbf{X}$ má lineárně nezávislé sloupce, právě když je $\mathbf{X}\boldsymbol{s} = \mathbf{0}$ pouze pro $\boldsymbol{s} = \mathbf{0}$.

- Obecně pro matici $\mathbf{X}$ a libovolný vektor $\boldsymbol{s} \in \mathbb{R}^{p+1}$ platí

$$\mathbf{X}\boldsymbol{s} = \mathbf{0} \;\Rightarrow\; \mathbf{X}^T\mathbf{X}\boldsymbol{s} = \mathbf{0} \;\Rightarrow\; \boldsymbol{s}^T\mathbf{X}^T\mathbf{X}\boldsymbol{s} = 0 \;\Rightarrow\; \|\mathbf{X}\boldsymbol{s}\|^2 = 0 \;\Rightarrow\; \mathbf{X}\boldsymbol{s} = \mathbf{0}.$$

- Z toho plyne, že je $\mathbf{X}^T\mathbf{X}$ je regulární, právě když jsou sloupce matice $\mathbf{X}$ lineárně nezávislé.

Regularita versus lineární nezávislost sloupců (2/3)

- Problém zcela určitě nastává, pokud $N < p + 1$. Pak totiž v N rozměrném prostoru $\mathbb{R}^N$ nemůže existovat $p + 1$ lineárně nezávislých vektorů a tak ani sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ matice $\mathbf{X}$ nemohou být lineárně nezávislé.

Regularita versus lineární nezávislost sloupců (2/3)

- Problém zcela určitě nastává, pokud $N < p + 1$. Pak totiž v N rozměrném prostoru $\mathbb{R}^N$ nemůže existovat $p + 1$ lineárně nezávislých vektorů a tak ani sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ matice $\mathbf{X}$ nemohou být lineárně nezávislé.

- I v situaci $N \geq p + 1$ se ale může stát, že sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ nebudou lineárně nezávislé.

Regularita versus lineární nezávislost sloupců (2/3)

- Problém zcela určitě nastává, pokud $N < p + 1$. Pak totiž v N rozměrném prostoru $\mathbb{R}^N$ nemůže existovat $p + 1$ lineárně nezávislých vektorů a tak ani sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ matice $\mathbf{X}$ nemohou být lineárně nezávislé.

- I v situaci $N \geq p + 1$ se ale může stát, že sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ nebudou lineárně nezávislé.

- Může to být například tím, že přímo jednotlivé příznaky jsou lineárně závislé a tedy jeden z nich je lineární kombinací ostatních.

Regularita versus lineární nezávislost sloupců (2/3)

- Problém zcela určitě nastává, pokud $N < p + 1$. Pak totiž v N rozměrném prostoru $\mathbb{R}^N$ nemůže existovat $p + 1$ lineárně nezávislých vektorů a tak ani sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ matice $\mathbf{X}$ nemohou být lineárně nezávislé.

- I v situaci $N \geq p + 1$ se ale může stát, že sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ nebudou lineárně nezávislé.

- Může to být například tím, že přímo jednotlivé příznaky jsou lineárně závislé a tedy jeden z nich je lineární kombinací ostatních.

- V takovém případě nepomůže ani libovolně vysoké N a sloupce matice $\mathbf{X}$ budou lineárně závislé vždy.

Regularita versus lineární nezávislost sloupců (2/3)

- Problém zcela určitě nastává, pokud $N < p + 1$. Pak totiž v N rozměrném prostoru $\mathbb{R}^N$ nemůže existovat $p + 1$ lineárně nezávislých vektorů a tak ani sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ matice $\mathbf{X}$ nemohou být lineárně nezávislé.

- I v situaci $N \geq p + 1$ se ale může stát, že sloupce $\mathbf{X}_{\bullet 0}, \ldots, \mathbf{X}_{\bullet p}$ nebudou lineárně nezávislé.

- Může to být například tím, že přímo jednotlivé příznaky jsou lineárně závislé a tedy jeden z nich je lineární kombinací ostatních.

- V takovém případě nepomůže ani libovolně vysoké N a sloupce matice $\mathbf{X}$ budou lineárně závislé vždy.

- Podívejme se, co to znamená pro řešení úlohy minimalizace $\mathrm{RSS}(\boldsymbol{w})$.

Regularita versus lineární nezávislost sloupců (3/3)

- Normální rovnice $\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}$ je z pohledu složek vektoru $\boldsymbol{w}$ soustava $p+1$ lineárních rovnic o $p+1$ neznámých.

Regularita versus lineární nezávislost sloupců (3/3)

- Normální rovnice $\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}$ je z pohledu složek vektoru $\boldsymbol{w}$ soustava $p+1$ lineárních rovnic o $p+1$ neznámých.
- Tato soustava má vždy alespoň jedno řešení. Pokud jsou sloupce matice $\mathbf{X}$ lineárně nezávislé, je $\mathbf{X}^T\mathbf{X}$ regulární, řešení je právě jedno a to

$$\hat{\boldsymbol{w}}_{\mathsf{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}.$$

Regularita versus lineární nezávislost sloupců (3/3)

- Normální rovnice $\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}$ je z pohledu složek vektoru $\boldsymbol{w}$ soustava $p+1$ lineárních rovnic o $p+1$ neznámých.
- Tato soustava má vždy alespoň jedno řešení. Pokud jsou sloupce matice $\mathbf{X}$ lineárně nezávislé, je $\mathbf{X}^T\mathbf{X}$ regulární, řešení je právě jedno a to

$$\hat{\boldsymbol{w}}_{\mathsf{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}.$$

- V opačném případě existuje nekonečně mnoho řešení tak, že pro každé dvě řešení $\boldsymbol{w}$ a $\boldsymbol{w}'$ platí $\mathbf{X}(\boldsymbol{w} - \boldsymbol{w}') = \mathbf{0}$.

Regularita versus lineární nezávislost sloupců (3/3)

- Normální rovnice $\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}$ je z pohledu složek vektoru $\boldsymbol{w}$ soustava $p+1$ lineárních rovnic o $p+1$ neznámých.
- Tato soustava má vždy alespoň jedno řešení. Pokud jsou sloupce matice $\mathbf{X}$ lineárně nezávislé, je $\mathbf{X}^T\mathbf{X}$ regulární, řešení je právě jedno a to

$$\hat{\boldsymbol{w}}_{\mathsf{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}.$$

- V opačném případě existuje nekonečně mnoho řešení tak, že pro každé dvě řešení $\boldsymbol{w}$ a $\boldsymbol{w}'$ platí $\mathbf{X}(\boldsymbol{w} - \boldsymbol{w}') = \mathbf{0}$.
- Poznamenejme, že pro každou dvojici řešení platí

$$\begin{aligned}\mathrm{RSS}(\boldsymbol{w}) &= \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w} + \mathbf{X}\boldsymbol{w}' - \mathbf{X}\boldsymbol{w}'\|^2 \\ &= \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}' - \mathbf{X}(\boldsymbol{w} - \boldsymbol{w}')\|^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}'\|^2 = \mathrm{RSS}(\boldsymbol{w}').\end{aligned}$$

Regularita versus lineární nezávislost sloupců (3/3)

- Normální rovnice $\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{0}$ je z pohledu složek vektoru $\boldsymbol{w}$ soustava $p+1$ lineárních rovnic o $p+1$ neznámých.
- Tato soustava má vždy alespoň jedno řešení. Pokud jsou sloupce matice $\mathbf{X}$ lineárně nezávislé, je $\mathbf{X}^T\mathbf{X}$ regulární, řešení je právě jedno a to

$$\hat{\boldsymbol{w}}_{\mathrm{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}.$$

- V opačném případě existuje nekonečně mnoho řešení tak, že pro každé dvě řešení $\boldsymbol{w}$ a $\boldsymbol{w}'$ platí $\mathbf{X}(\boldsymbol{w} - \boldsymbol{w}') = \mathbf{0}$.
- Poznamenejme, že pro každou dvojici řešení platí

$$\begin{aligned}\mathrm{RSS}(\boldsymbol{w}) &= \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w} + \mathbf{X}\boldsymbol{w}' - \mathbf{X}\boldsymbol{w}'\|^2 \\ &= \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}' - \mathbf{X}(\boldsymbol{w} - \boldsymbol{w}')\|^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}'\|^2 = \mathrm{RSS}(\boldsymbol{w}').\end{aligned}$$

- Všechny řešení tedy odpovídají stejné hodnotě RSS, která je, jak jsme již zmínili, globálním minimem, které je v tomto případě neostré.

Řešení při lineární závislých sloupcích

- Otázkou je, jak nějaké řešení získat, když nemůžeme invertovat matici $\mathbf{X}^T\mathbf{X}$ a následně použít vzoreček $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$.

Řešení při lineární závislých sloupcích

- Otázkou je, jak nějaké řešení získat, když nemůžeme invertovat matici $\mathbf{X}^T\mathbf{X}$ a následně použít vzoreček $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$.
- Funkce `LinearRegression` z balíčku `scikit-learn` si s takovými případy poradí a řešení vrátí.[1]

[1] Jen pro zajímavost uveďme, že uvnitř se využívá funkce `dgelsd` z knihovny LAPACK.

Řešení při lineární závislých sloupcích

- Otázkou je, jak nějaké řešení získat, když nemůžeme invertovat matici $\mathbf{X}^T\mathbf{X}$ a následně použít vzoreček $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$.
- Funkce `LinearRegression` z balíčku `scikit-learn` si s takovými případy poradí a řešení vrátí.[1]
- Pokud $\mathbf{X}^T\mathbf{X}$ není regulární, vrátí takový vektor $\hat{\boldsymbol{w}}$, který řeší normální rovnici a zároveň má mezi všemi řešeními nejmenší normu $\|\hat{\boldsymbol{w}}\|$.

[1] Jen pro zajímavost uveďme, že uvnitř se využívá funkce `dgelsd` z knihovny LAPACK.

Řešení při lineární závislých sloupcích

- Otázkou je, jak nějaké řešení získat, když nemůžeme invertovat matici $\mathbf{X}^T\mathbf{X}$ a následně použít vzoreček $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$.
- Funkce `LinearRegression` z balíčku `scikit-learn` si s takovými případy poradí a řešení vrátí.[1]
- Pokud $\mathbf{X}^T\mathbf{X}$ není regulární, vrátí takový vektor $\hat{\boldsymbol{w}}$, který řeší normální rovnici a zároveň má mezi všemi řešeními nejmenší normu $\|\hat{\boldsymbol{w}}\|$.
- Jen pro zajímavost uveďme, že toto řešení lze zapsat ve tvaru

$$\hat{\boldsymbol{w}} = (\mathbf{X}^T\mathbf{X})^{+}\mathbf{X}^T\boldsymbol{Y},$$

kde $(\mathbf{X}^T\mathbf{X})^{+}$ je takzvaná Moorova-Penroseova pseudoinverzní matice k matici $\mathbf{X}^T\mathbf{X}$.

[1] Jen pro zajímavost uveďme, že uvnitř se využívá funkce `dgelsd` z knihovny LAPACK.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

- V takovém případě sice inverze $\mathbf{X}^T\mathbf{X}$ teoreticky existuje, ale prakticky je její výpočet numericky problematický.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

- V takovém případě sice inverze $\mathbf{X}^T\mathbf{X}$ teoreticky existuje, ale prakticky je její výpočet numericky problematický.
- Především - a to je to hlavní jádro problému - je získaný odhad $\hat{\boldsymbol{w}}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$ velmi citlivý na malé nevhodné změny $\boldsymbol{Y}$.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

- V takovém případě sice inverze $\mathbf{X}^T\mathbf{X}$ teoreticky existuje, ale prakticky je její výpočet numericky problematický.
- Především - a to je to hlavní jádro problému - je získaný odhad $\hat{\boldsymbol{w}}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$ velmi citlivý na malé nevhodné změny $\boldsymbol{Y}$.
- To znamená, že kdybychom náhodný výběr trénovací množiny zopakovali, může se hodnota odhadu $\hat{\boldsymbol{w}}_{\text{OLS}}$ radikálně změnit.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

- V takovém případě sice inverze $\mathbf{X}^T\mathbf{X}$ teoreticky existuje, ale prakticky je její výpočet numericky problematický.
- Především - a to je to hlavní jádro problému - je získaný odhad $\hat{\boldsymbol{w}}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$ velmi citlivý na malé nevhodné změny $\boldsymbol{Y}$.
- To znamená, že kdybychom náhodný výběr trénovací množiny zopakovali, může se hodnota odhadu $\hat{\boldsymbol{w}}_{\text{OLS}}$ radikálně změnit.
- Z pravděpodobnostního pohledu lze ukázat, že $\hat{\boldsymbol{w}}_{\text{OLS}}$ má potom v jistých směrech velký rozptyl.

Problém kolinearity

- Problémem nejsou pouze případy, kdy jsou sloupce matice $\boldsymbol{X}$ lineárně závislé, ale úplně stačí, když jsou „skoro“ lineárně závislé.
- V obou těchto případech mluvíme o problému **kolinearity** (angl. **collinearity**).
- Myslíme tím tedy, že existují lineární kombinace sloupců, které dávají téměř nulové vektory, zatímco jiné lineární kombinace vrací mnohem větší vektory, tj.

$$\|\mathbf{X}\boldsymbol{u}\| \gg \|\mathbf{X}\boldsymbol{v}\| \doteq 0 \quad \text{pro nějaké} \quad \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1.$$

- V takovém případě sice inverze $\mathbf{X}^T\mathbf{X}$ teoreticky existuje, ale prakticky je její výpočet numericky problematický.
- Především - a to je to hlavní jádro problému - je získaný odhad $\hat{\boldsymbol{w}}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}$ velmi citlivý na malé nevhodné změny $\boldsymbol{Y}$.
- To znamená, že kdybychom náhodný výběr trénovací množiny zopakovali, může se hodnota odhadu $\hat{\boldsymbol{w}}_{\text{OLS}}$ radikálně změnit.
- Z pravděpodobnostního pohledu lze ukázat, že $\hat{\boldsymbol{w}}_{\text{OLS}}$ má potom v jistých směrech velký rozptyl.
- Toto se pochopitelně přenáší i na predikce $\hat{Y}$, které pak mají v některých bodech velký rozptyl, což znamená, že jim nemůžeme příliš důvěřovat.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

  To se ale nestane, pokud jsou samotné příznaky (skoro) lineárně závislé.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

  To se ale nestane, pokud jsou samotné příznaky (skoro) lineárně závislé.

- Snížit počet příznaků. To znamená vyhození některých příznaků, případně nahrazení příznaků menším počtem nových, které již nebudou lineárně závislé.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

  To se ale nestane, pokud jsou samotné příznaky (skoro) lineárně závislé.

- Snížit počet příznaků. To znamená vyhození některých příznaků, případně nahrazení příznaků menším počtem nových, které již nebudou lineárně závislé.

  Jednou z metod redukce počtu příznaků, kdy ty existující nahrazujeme menším počtem jejich lineárních kombinací, se budeme zabývat v 10. přednášce.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

  To se ale nestane, pokud jsou samotné příznaky (skoro) lineárně závislé.

- Snížit počet příznaků. To znamená vyhození některých příznaků, případně nahrazení příznaků menším počtem nových, které již nebudou lineárně závislé.

  Jednou z metod redukce počtu příznaků, kdy ty existující nahrazujeme menším počtem jejich lineárních kombinací, se budeme zabývat v 10. přednášce.

- Změnit funkci, kterou minimalizujeme, abychom měli jednoznačné a stabilní řešení.

Regularizace lineární regrese

Pokud narazíme na problém kolinearity, máme v zásadě tři možnosti:

- Přigenerovat další data nebo odebrat existující a doufat, že se problém vyřeší.

  To se ale nestane, pokud jsou samotné příznaky (skoro) lineárně závislé.

- Snížit počet příznaků. To znamená vyhození některých příznaků, případně nahrazení příznaků menším počtem nových, které již nebudou lineárně závislé.

  Jednou z metod redukce počtu příznaků, kdy ty existující nahrazujeme menším počtem jejich lineárních kombinací, se budeme zabývat v 10. přednášce.

- Změnit funkci, kterou minimalizujeme, abychom měli jednoznačné a stabilní řešení.

  Typicky provedeme tak zvanou regularizaci, kdy k RSS přidáme **regularizační člen**, který problémy kolinearity odstraní nebo alespoň dostatečně zmírní.

Hřebenová regrese (1/3)

Hřebenová regrese (angl. **ridge regression**) [Hoerl, Kennard (1970)] nebo taky L_2 regularizace se k problému kolinearity staví zavedením penalizačního členu úměrného kvadrátu normy vektoru koeficientů $\boldsymbol{w}$ bez interceptu.

Hřebenová regrese (1/3)

Hřebenová regrese (angl. **ridge regression**) [Hoerl, Kennard (1970)] nebo taky L_2 regularizace se k problému kolinearity staví zavedením penalizačního členu úměrného kvadrátu normy vektoru koeficientů $\boldsymbol{w}$ bez interceptu.

Minimalizujeme tedy **regularizovaný reziduální součet čtverců**

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2,$$

který závisí na parametru $\lambda \geq 0$.

Hřebenová regrese (1/3)

Hřebenová regrese (angl. **ridge regression**) [Hoerl, Kennard (1970)] nebo taky L_2 regularizace se k problému kolinearity staví zavedením penalizačního členu úměrného kvadrátu normy vektoru koeficientů $\boldsymbol{w}$ bez interceptu.

Minimalizujeme tedy **regularizovaný reziduální součet čtverců**

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2,$$

který závisí na parametru $\lambda \geq 0$.

- Pro $\lambda = 0$ dostáváme $\mathrm{RSS}_0(\boldsymbol{w}) = \mathrm{RSS}(\boldsymbol{w})$ a máme tedy obyčejnou metodu nejmenších čtverců.

Hřebenová regrese (1/3)

Hřebenová regrese (angl. **ridge regression**) [Hoerl, Kennard (1970)] nebo taky L_2 regularizace se k problému kolinearity staví zavedením penalizačního členu úměrného kvadrátu normy vektoru koeficientů $\boldsymbol{w}$ bez interceptu.

Minimalizujeme tedy **regularizovaný reziduální součet čtverců**

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2,$$

který závisí na parametru $\lambda \geq 0$.

- Pro $\lambda = 0$ dostáváme $\mathrm{RSS}_0(\boldsymbol{w}) = \mathrm{RSS}(\boldsymbol{w})$ a máme tedy obyčejnou metodu nejmenších čtverců.
- Pro $\lambda > 0$ je vidět, že v minimu se bude cílit na takové vektory $\boldsymbol{w}$, které mají co nejmenší složky.

Hřebenová regrese (1/3)

Hřebenová regrese (angl. **ridge regression**) [Hoerl, Kennard (1970)] nebo taky L_2 regularizace se k problému kolinearity staví zavedením penalizačního členu úměrného kvadrátu normy vektoru koeficientů $\boldsymbol{w}$ bez interceptu.

Minimalizujeme tedy **regularizovaný reziduální součet čtverců**

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2,$$

který závisí na parametru $\lambda \geq 0$.

- Pro $\lambda = 0$ dostáváme $\mathrm{RSS}_0(\boldsymbol{w}) = \mathrm{RSS}(\boldsymbol{w})$ a máme tedy obyčejnou metodu nejmenších čtverců.
- Pro $\lambda > 0$ je vidět, že v minimu se bude cílit na takové vektory $\boldsymbol{w}$, které mají co nejmenší složky.
- Hodnotu w_0 interceptu nijak nepenalizujeme. Jedná se pouze o vertikální posun, který zajišťuje předpoklad $\mathrm{E}\,\varepsilon = 0$ modelu a je tedy vhodné ho neomezovat.

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

můžeme psát

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

můžeme psát

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Gradient a Hessova matice jsou

$$\nabla \mathrm{RSS}_\lambda(\boldsymbol{w}) = -2\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) + 2\lambda \mathbf{I}' \boldsymbol{w}$$

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

můžeme psát

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Gradient a Hessova matice jsou

$$\nabla \mathrm{RSS}_\lambda(\boldsymbol{w}) = -2\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) + 2\lambda \mathbf{I}'\boldsymbol{w} \quad \text{a} \quad \mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}'.$$

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

můžeme psát

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Gradient a Hessova matice jsou

$$\nabla \mathrm{RSS}_\lambda(\boldsymbol{w}) = -2\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) + 2\lambda \mathbf{I}'\boldsymbol{w} \quad \text{a} \quad \mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}'.$$

Ekvivalent normální rovnice je tedy

$$\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} - \lambda\mathbf{I}'\boldsymbol{w} = \mathbf{0}.$$

Hřebenová regrese (2/3)

Zavedeme-li matici

$$\mathbf{I}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{p+1,p+1},$$

můžeme psát

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{p} w_i^2 = \|\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Gradient a Hessova matice jsou

$$\nabla \mathrm{RSS}_\lambda(\boldsymbol{w}) = -2\mathbf{X}^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{w}) + 2\lambda \mathbf{I}'\boldsymbol{w} \quad \text{a} \quad \mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}'.$$

Ekvivalent normální rovnice je tedy

$$\mathbf{X}^T\boldsymbol{Y} - \mathbf{X}^T\mathbf{X}\boldsymbol{w} - \lambda\mathbf{I}'\boldsymbol{w} = \mathbf{0}.$$

Jak za chvilku ukážeme, matice $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}'$ je pro $\lambda > 0$ regulární, a řešení je tedy jednoznačně

$$\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}.$$

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s}$$

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} = \|\mathbf{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,$$

protože pro $\boldsymbol{s} = (s_0, 0, \ldots, 0)^T \neq \mathbf{0}$ máme $\mathbf{X}\boldsymbol{s} = (s_0, \ldots, s_0)^T \neq \mathbf{0}$.

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} = \|\mathbf{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,$$

protože pro $\boldsymbol{s} = (s_0, 0, \ldots, 0)^T \neq \mathbf{0}$ máme $\mathbf{X}\boldsymbol{s} = (s_0, \ldots, s_0)^T \neq \mathbf{0}$.

Hessova matice je tedy pozitivně definitní a matice $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}'$ je regulární.

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} = \|\mathbf{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,$$

protože pro $\boldsymbol{s} = (s_0, 0, \ldots, 0)^T \neq \mathbf{0}$ máme $\mathbf{X}\boldsymbol{s} = (s_0, \ldots, s_0)^T \neq \mathbf{0}$.

Hessova matice je tedy pozitivně definitní a matice $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}'$ je regulární.

Pro $\lambda > 0$ tak vždy existuje jednoznačné řešení

$$\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$$

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} = \|\mathbf{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,$$

protože pro $\boldsymbol{s} = (s_0, 0, \ldots, 0)^T \neq \mathbf{0}$ máme $\mathbf{X}\boldsymbol{s} = (s_0, \ldots, s_0)^T \neq \mathbf{0}$.

Hessova matice je tedy pozitivně definitní a matice $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}'$ je regulární.

Pro $\lambda > 0$ tak vždy existuje jednoznačné řešení

$$\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$$

a odpovídá globálnímu minimu RSS_λ.

Hřebenová regrese (3/3)

Podle předchozího slajdu je Hessova matice

$$\mathbf{H}_{\mathrm{RSS}_\lambda}(\boldsymbol{w}) = 2\mathbf{X}^T\mathbf{X} + 2\lambda\mathbf{I}' = 2(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}').$$

Dále pro každé $\boldsymbol{s} \in \mathbb{R}^{p+1}, \boldsymbol{s} \neq \mathbf{0}$ a $\lambda > 0$ platí

$$\boldsymbol{s}^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')\boldsymbol{s} = (\mathbf{X}\boldsymbol{s})^T(\mathbf{X}\boldsymbol{s}) + \lambda\boldsymbol{s}^T\mathbf{I}'\boldsymbol{s} = \|\mathbf{X}\boldsymbol{s}\|^2 + \lambda\sum_{i=1}^{p} s_i^2 > 0,$$

protože pro $\boldsymbol{s} = (s_0, 0, \ldots, 0)^T \neq \mathbf{0}$ máme $\mathbf{X}\boldsymbol{s} = (s_0, \ldots, s_0)^T \neq \mathbf{0}$.

Hessova matice je tedy pozitivně definitní a matice $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}'$ je regulární.

Pro $\lambda > 0$ tak vždy existuje jednoznačné řešení

$$\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$$

a odpovídá globálnímu minimu RSS_λ.

Predikce v bodě $\boldsymbol{x}$ je potom opět $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) = \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right)$$

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\begin{aligned}\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) &= \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right)\\ &= (\mathrm{E}\,Y)^2 - \mathrm{E}(Y\hat{Y}) - (\mathrm{E}\,Y)^2 + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y}\end{aligned}$$

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\begin{aligned}
\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) &= \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right) \\
&= (\mathrm{E}\,Y)^2 - \mathrm{E}(Y\hat{Y}) - (\mathrm{E}\,Y)^2 + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y} \\
&= -\mathrm{E}(Y\hat{Y}) + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y} = 0.
\end{aligned}$$

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\begin{aligned}\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) &= \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right)\\ &= (\mathrm{E}\,Y)^2 - \mathrm{E}(Y\hat{Y}) - (\mathrm{E}\,Y)^2 + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y}\\ &= -\mathrm{E}(Y\hat{Y}) + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y} = 0.\end{aligned}$$

Pro očekávanou chybu tedy platí

$$\mathrm{E}\,L(Y, \hat{Y}) = \mathrm{E}(Y - \hat{Y})^2$$

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\begin{aligned}\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) &= \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right)\\ &= (\mathrm{E}\,Y)^2 - \mathrm{E}(Y\hat{Y}) - (\mathrm{E}\,Y)^2 + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y}\\ &= -\mathrm{E}(Y\hat{Y}) + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y} = 0.\end{aligned}$$

Pro očekávanou chybu tedy platí

$$\mathrm{E}\,L(Y, \hat{Y}) = \mathrm{E}(Y - \hat{Y})^2 = \mathrm{E}(Y - \mathrm{E}\,Y + \mathrm{E}\,Y - \hat{Y})^2$$

Očekávaná chyba modelu

Jelikož $\boldsymbol{Y} = \mathbf{X}\boldsymbol{w} + \boldsymbol{\varepsilon}$ z trénovací množiny je v důsledku náhodnosti $\boldsymbol{\varepsilon}$ náhodný vektor, dostáváme, že i $\hat{\boldsymbol{w}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}')^{-1}\mathbf{X}^T\boldsymbol{Y}$ je jakožto funkce $\boldsymbol{Y}$ náhodný vektor.

Uvažujme nějaký pevný bod $\boldsymbol{x} = (1, x_1, \ldots, x_p) \in \mathbb{R}^{p+1}$ a zkoumejme očekávanou chybu měřenou pomocí kvadratické ztrátové funkce při odhadu $Y = \boldsymbol{x}^T\boldsymbol{w} + \varepsilon$ pomocí $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{w}}_\lambda$.

Budeme předpokládat nezávislost trénovacích a testovacích dat, tj. nezávislost $\boldsymbol{Y}$ a Y, a v důsledku tedy nezávislost $\hat{Y}$ a Y.

Z toho plyne

$$\begin{aligned}\mathrm{E}\left((Y - \mathrm{E}\,Y)(\mathrm{E}\,Y - \hat{Y})\right) &= \mathrm{E}\left(Y(\mathrm{E}\,Y) - (Y\hat{Y}) - (\mathrm{E}\,Y)^2 + (\mathrm{E}\,Y)\hat{Y}\right)\\ &= (\mathrm{E}\,Y)^2 - \mathrm{E}(Y\hat{Y}) - (\mathrm{E}\,Y)^2 + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y}\\ &= -\mathrm{E}(Y\hat{Y}) + \mathrm{E}\,Y\,\mathrm{E}\,\hat{Y} = 0.\end{aligned}$$

Pro očekávanou chybu tedy platí

$$\mathrm{E}\,L(Y, \hat{Y}) = \mathrm{E}(Y - \hat{Y})^2 = \mathrm{E}(Y - \mathrm{E}\,Y + \mathrm{E}\,Y - \hat{Y})^2 = \mathrm{E}(Y - \mathrm{E}\,Y)^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\,Y)^2.$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\operatorname{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\operatorname{E} Y$ (angl. **mean squared error**).

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\mathrm{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\operatorname{E} Y$ (angl. **mean squared error**).

$$\mathrm{MSE}(\hat{Y}) = \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2 = \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y + \hat{Y} - \operatorname{E} \hat{Y})^2$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\mathrm{E}\, L(Y, \hat{Y}) = \sigma^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\mathrm{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\mathrm{E}\, Y$ (angl. **mean squared error**).

$$\begin{aligned}\mathrm{MSE}(\hat{Y}) &= \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2 = \mathrm{E}(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y + \hat{Y} - \mathrm{E}\, \hat{Y})^2 \\ &= \mathrm{E}(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})^2 + 2\,\mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)\end{aligned}$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\mathrm{E}\, L(Y, \hat{Y}) = \sigma^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\mathrm{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\mathrm{E}\, Y$ (angl. **mean squared error**).

$$\begin{aligned}
\mathrm{MSE}(\hat{Y}) &= \mathrm{E}(\hat{Y} - \mathrm{E}\, Y)^2 = \mathrm{E}(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y + \hat{Y} - \mathrm{E}\, \hat{Y})^2 \\
&= \mathrm{E}(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})^2 + 2\, \mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})(\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y) \\
&= (\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)^2 + \mathrm{E}(\hat{Y} - \mathrm{E}\, \hat{Y})^2 + 2 \cdot 0 \cdot (\mathrm{E}\, \hat{Y} - \mathrm{E}\, Y)
\end{aligned}$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\operatorname{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\operatorname{E} Y$ (angl. **mean squared error**).

$$\begin{aligned}
\operatorname{MSE}(\hat{Y}) &= \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2 = \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y + \hat{Y} - \operatorname{E} \hat{Y})^2 \\
&= \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})(\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \cdot 0 \cdot (\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{var} \hat{Y} = (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y},
\end{aligned}$$

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\operatorname{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\operatorname{E} Y$ (angl. **mean squared error**).

$$\begin{aligned}
\operatorname{MSE}(\hat{Y}) &= \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2 = \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y + \hat{Y} - \operatorname{E} \hat{Y})^2 \\
&= \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})(\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \cdot 0 \cdot (\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{var} \hat{Y} = (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y},
\end{aligned}$$

kde $\operatorname{bias} \hat{Y} = \operatorname{E} \hat{Y} - \operatorname{E} Y$ značí vychýlení odhadu (angl. bias).

Rozklad očekávané chyby modelu

Označíme-li $\operatorname{var} Y = \operatorname{var} \varepsilon = \sigma^2$ dostáváme

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2.$$

První člen odpovídá neodstranitelné chybě, která je dána náhodností v modelu. Tato chyba se nazývá Bayesovská (**Bayes error**).
Druhý člen se značí $\operatorname{MSE}(\hat{Y})$ a nazývá střední kvadratická chyba odhadu $\hat{Y}$ parametru $\operatorname{E} Y$ (angl. **mean squared error**).

$$\begin{aligned}
\operatorname{MSE}(\hat{Y}) &= \operatorname{E}(\hat{Y} - \operatorname{E} Y)^2 = \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y + \hat{Y} - \operatorname{E} \hat{Y})^2 \\
&= \operatorname{E}(\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})(\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{E}(\hat{Y} - \operatorname{E} \hat{Y})^2 + 2 \cdot 0 \cdot (\operatorname{E} \hat{Y} - \operatorname{E} Y) \\
&= (\operatorname{E} \hat{Y} - \operatorname{E} Y)^2 + \operatorname{var} \hat{Y} = (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y},
\end{aligned}$$

kde $\operatorname{bias} \hat{Y} = \operatorname{E} \hat{Y} - \operatorname{E} Y$ značí vychýlení odhadu (angl. bias).
Dohromady tedy máme finální dekompozici očekávané chyby jako

$$\operatorname{E} L(Y, \hat{Y}) = \sigma^2 + (\operatorname{bias} \hat{Y})^2 + \operatorname{var} \hat{Y}.$$

Bias-variance tradeoff

U hřebenové regrese lze ukázat, že (hodně zjednodušeně) platí

$$(\operatorname{bias} \hat{Y})^2 \sim \left(1 - \frac{1}{1+\lambda}\right)^2 \quad \text{a} \quad \operatorname{var} \hat{Y} \sim \left(\frac{1}{1+\lambda}\right)^2.$$

Bias-variance tradeoff

U hřebenové regrese lze ukázat, že (hodně zjednodušeně) platí

$$(\operatorname{bias} \hat{Y})^2 \sim \left(1 - \frac{1}{1+\lambda}\right)^2 \quad \text{a} \quad \operatorname{var} \hat{Y} \sim \left(\frac{1}{1+\lambda}\right)^2.$$

To znamená, že s rostoucím λ vychýlení roste a rozptyl klesá.

Bias-variance tradeoff

U hřebenové regrese lze ukázat, že (hodně zjednodušeně) platí

$$(\operatorname{bias} \hat{Y})^2 \sim \left(1 - \frac{1}{1+\lambda}\right)^2 \quad \text{a} \quad \operatorname{var} \hat{Y} \sim \left(\frac{1}{1+\lambda}\right)^2.$$

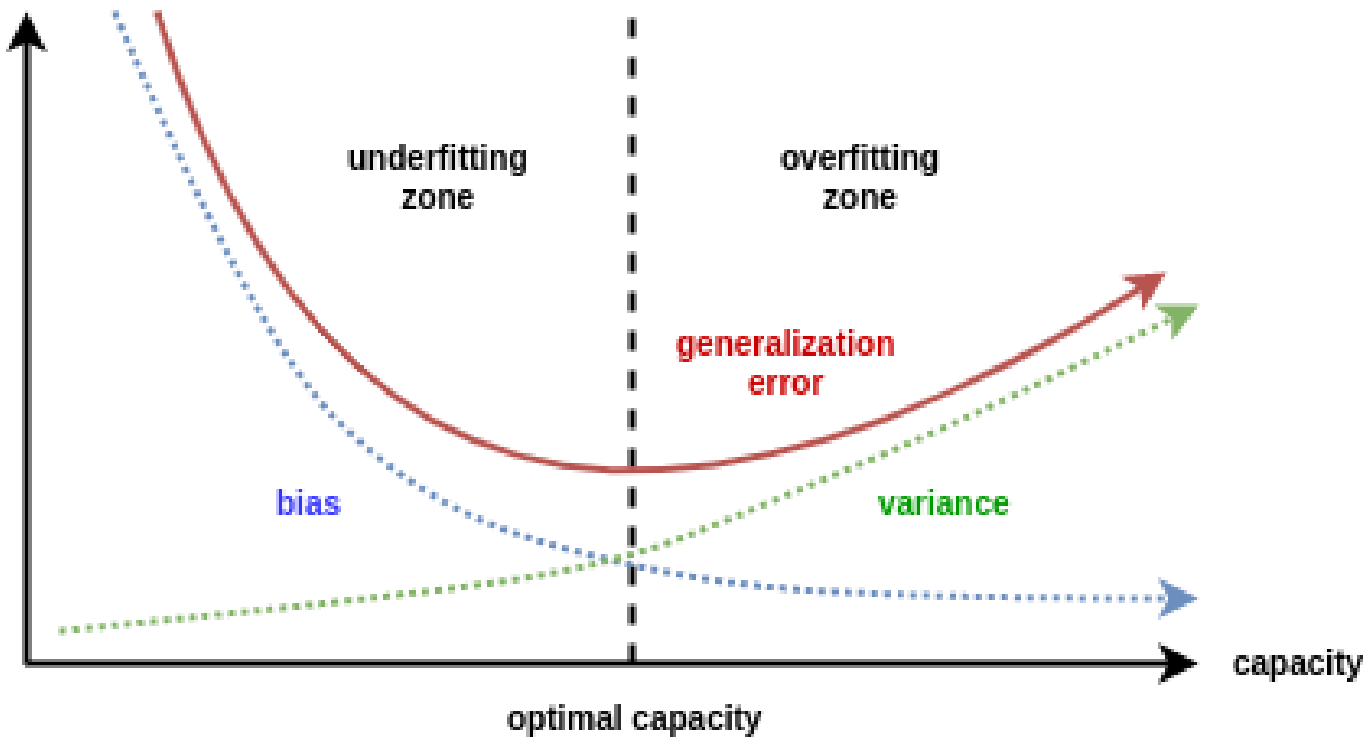
To znamená, že s rostoucím λ vychýlení roste a rozptyl klesá. Takovéto chování v závislosti na hyperparametrech modelu je typické a nazývá se **bias-variance tradeoff**.

Bias-variance tradeoff

U hřebenové regrese lze ukázat, že (hodně zjednodušeně) platí

$$(\operatorname{bias} \hat{Y})^2 \sim \left(1 - \frac{1}{1+\lambda}\right)^2 \quad \text{a} \quad \operatorname{var} \hat{Y} \sim \left(\frac{1}{1+\lambda}\right)^2.$$

To znamená, že s rostoucím λ vychýlení roste a rozptyl klesá. Takovéto chování v závislosti na hyperparametrech modelu je typické a nazývá se **bias-variance tradeoff**.



Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.

Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.
- Obvykle se snažíme minimalizovat odhad MSE validační množiny dat případně odhad MSE pomocí cross-validace.

Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.
- Obvykle se snažíme minimalizovat odhad MSE validační množiny dat případně odhad MSE pomocí cross-validace. Odhad MSE se pro validační množinu $(Y'_i, \boldsymbol{x}'_i)$ velikosti n počítá jako

$$\mathsf{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y'_i - \boldsymbol{x}'^T_i \hat{\boldsymbol{w}}_\lambda\right)^2.$$

Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.
- Obvykle se snažíme minimalizovat odhad MSE validační množiny dat případně odhad MSE pomocí cross-validace. Odhad MSE se pro validační množinu $(Y'_i, \boldsymbol{x}'_i)$ velikosti n počítá jako

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y'_i - \boldsymbol{x}'^T_i \hat{\boldsymbol{w}}_\lambda\right)^2.$$

- Při použití hřebenové regrese bývá obvyklé nejprve jednotlivé příznaky standardizovat, aby se staly rozsahově porovnatelné a tedy, aby byly penalizovány všechny stejně.

Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.
- Obvykle se snažíme minimalizovat odhad MSE validační množiny dat případně odhad MSE pomocí cross-validace. Odhad MSE se pro validační množinu $(Y'_i, \boldsymbol{x}'_i)$ velikosti n počítá jako

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y'_i - \boldsymbol{x}'^T_i \hat{\boldsymbol{w}}_\lambda\right)^2.$$

- Při použití hřebenové regrese bývá obvyklé nejprve jednotlivé příznaky standardizovat, aby se staly rozsahově porovnatelné a tedy, aby byly penalizovány všechny stejně. Tj. místo příznaku X_i použijeme příznak

$$X'_i = \frac{X_i - \bar{X}_i}{\sqrt{s^2_{X_i}}}, \quad \text{kde} \quad \bar{X}_i = \frac{1}{N} \sum_{j=1}^{N} x_{j;i} \quad \text{a} \quad s^2_{X_i} = \frac{1}{N-1} \sum_{j=1}^{N} (x_{j;i} - \bar{X}_i)^2$$

Různé poznámky

- Hledáme tedy optimální hodnotu parametru λ, pro kterou je chyba modelu nejmenší.
- Obvykle se snažíme minimalizovat odhad MSE validační množiny dat případně odhad MSE pomocí cross-validace. Odhad MSE se pro validační množinu $(Y'_i, \boldsymbol{x}'_i)$ velikosti n počítá jako

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y'_i - \boldsymbol{x}'^{T}_i \hat{\boldsymbol{w}}_\lambda\right)^2.$$

- Při použití hřebenové regrese bývá obvyklé nejprve jednotlivé příznaky standardizovat, aby se staly rozsahově porovnatelné a tedy, aby byly penalizovány všechny stejně. Tj. místo příznaku X_i použijeme příznak

$$X'_i = \frac{X_i - \bar{X}_i}{\sqrt{s^2_{X_i}}}, \quad \text{kde} \quad \bar{X}_i = \tfrac{1}{N}\sum_{j=1}^{N} x_{j;i} \quad \text{a} \quad s^2_{X_i} = \tfrac{1}{N-1}\sum_{j=1}^{N}(x_{j;i} - \bar{X}_i)^2$$

- Existují i další možnosti regularizace jako např. $\lambda \sum_{i=1}^{p} |w_i|$ (Lasso).

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.
- Základní rozšíření spočívá v nahrazení původních příznaků jejich transformovanými variantami.

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.
- Základní rozšíření spočívá v nahrazení původních příznaků jejich transformovanými variantami.
- Pro $M \in \mathbb{N}$ vezměme M funkcí $\varphi_1, \ldots, \varphi_M$ z $\mathbb{R}^p$ do $\mathbb{R}$ reprezentujících transformace X a nazvěme je **bázové funkce** (angl. **basis functions**).

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.
- Základní rozšíření spočívá v nahrazení původních příznaků jejich transformovanými variantami.
- Pro $M \in \mathbb{N}$ vezměme M funkcí $\varphi_1, \ldots, \varphi_M$ z $\mathbb{R}^p$ do $\mathbb{R}$ reprezentujících transformace X a nazvěme je **bázové funkce** (angl. **basis functions**).
- K těmto funkcím přidáme $\phi_0(\boldsymbol{x}) = 1$ a poskládáme je do vektorové funkce $\boldsymbol{\varphi} : \mathbb{R}^p \to \mathbb{R}^{M+1}$ vztahem $\boldsymbol{\varphi}(\boldsymbol{x}) = (1, \varphi_1(\boldsymbol{x}), \ldots, \varphi_M(\boldsymbol{x}))^T$.

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.
- Základní rozšíření spočívá v nahrazení původních příznaků jejich transformovanými variantami.
- Pro $M \in \mathbb{N}$ vezměme M funkcí $\varphi_1, \ldots, \varphi_M$ z $\mathbb{R}^p$ do $\mathbb{R}$ reprezentujících transformace X a nazvěme je **bázové funkce** (angl. **basis functions**).
- K těmto funkcím přidáme $\phi_0(\boldsymbol{x}) = 1$ a poskládáme je do vektorové funkce $\boldsymbol{\varphi} : \mathbb{R}^p \to \mathbb{R}^{M+1}$ vztahem $\boldsymbol{\varphi}(\boldsymbol{x}) = (1, \varphi_1(\boldsymbol{x}), \ldots, \varphi_M(\boldsymbol{x}))^T$.
- Jako model vztahu Y a $\boldsymbol{x}$ budeme uvažovat lineární model

$$Y = \sum_{j=0}^{M} w_j \varphi_j(\boldsymbol{x}) + \varepsilon = \boldsymbol{\varphi}(\boldsymbol{x})^T \boldsymbol{w} + \varepsilon,$$

Modely bázových funkcí (1/2)

- Doposud jsme uvažovali pouze jednoduchý lineární model ve tvaru

$$Y = \boldsymbol{x}^T \boldsymbol{w} + \varepsilon.$$

- Principiálně jsme tak schopni modelovat pouze lineární funkci ve vstupních proměnných. Ukažme si, jak rozšířit naše možnosti za obzor linearity.
- Základní rozšíření spočívá v nahrazení původních příznaků jejich transformovanými variantami.
- Pro $M \in \mathbb{N}$ vezměme M funkcí $\varphi_1, \ldots, \varphi_M$ z $\mathbb{R}^p$ do $\mathbb{R}$ reprezentujících transformace X a nazvěme je **bázové funkce** (angl. **basis functions**).
- K těmto funkcím přidáme $\phi_0(\boldsymbol{x}) = 1$ a poskládáme je do vektorové funkce $\boldsymbol{\varphi} : \mathbb{R}^p \to \mathbb{R}^{M+1}$ vztahem $\boldsymbol{\varphi}(\boldsymbol{x}) = (1, \varphi_1(\boldsymbol{x}), \ldots, \varphi_M(\boldsymbol{x}))^T$.
- Jako model vztahu Y a $\boldsymbol{x}$ budeme uvažovat lineární model

$$Y = \sum_{j=0}^{M} w_j \varphi_j(\boldsymbol{x}) + \varepsilon = \boldsymbol{\varphi}(\boldsymbol{x})^T \boldsymbol{w} + \varepsilon,$$

- Další postup je nyní zcela analogický jako před tím.

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.
- Maticově můžeme zapsat model pro trénovací data jako $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.
- Maticově můžeme zapsat model pro trénovací data jako $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}(\boldsymbol{x}_1)^T \\ \vdots \\ \boldsymbol{\varphi}(\boldsymbol{x}_N)^T \end{pmatrix} = \begin{pmatrix} 1 & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_M(\boldsymbol{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_M(\boldsymbol{x}_N) \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.
- Maticově můžeme zapsat model pro trénovací data jako $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}(\boldsymbol{x}_1)^T \\ \vdots \\ \boldsymbol{\varphi}(\boldsymbol{x}_N)^T \end{pmatrix} = \begin{pmatrix} 1 & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_M(\boldsymbol{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_M(\boldsymbol{x}_N) \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Obecně budeme minimalizovat (pro $\lambda = 0$ máme metodu nejmenších čtverců)

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.
- Maticově můžeme zapsat model pro trénovací data jako $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}(\boldsymbol{x}_1)^T \\ \vdots \\ \boldsymbol{\varphi}(\boldsymbol{x}_N)^T \end{pmatrix} = \begin{pmatrix} 1 & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_M(\boldsymbol{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_M(\boldsymbol{x}_N) \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Obecně budeme minimalizovat (pro $\lambda = 0$ máme metodu nejmenších čtverců)

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

- Řešením je (pro $\lambda = 0$ značíme také $\hat{\boldsymbol{w}}_{\mathrm{OLS}}$)

$$\hat{\boldsymbol{w}}_\lambda = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda \mathbf{I}'\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{Y}.$$

Modely bázových funkcí (2/2)

- Mějme tedy trénovací množinu jako náhodný výběr z výše uvedeného modelu určený N páry typu $(Y_i, \boldsymbol{x}_i)$.
- Maticově můžeme zapsat model pro trénovací data jako $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\varepsilon}$, kde

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}(\boldsymbol{x}_1)^T \\ \vdots \\ \boldsymbol{\varphi}(\boldsymbol{x}_N)^T \end{pmatrix} = \begin{pmatrix} 1 & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_M(\boldsymbol{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_M(\boldsymbol{x}_N) \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

- Obecně budeme minimalizovat (pro $\lambda = 0$ máme metodu nejmenších čtverců)

$$\mathrm{RSS}_\lambda(\boldsymbol{w}) = \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{w}\|^2 + \lambda \boldsymbol{w}^T \mathbf{I}' \boldsymbol{w}.$$

- Řešením je (pro $\lambda = 0$ značíme také $\hat{\boldsymbol{w}}_{\mathrm{OLS}}$)

$$\hat{\boldsymbol{w}}_\lambda = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda \mathbf{I}'\right)^{-1} \boldsymbol{\Phi}^T \boldsymbol{Y}.$$

- Predikce hodnoty Y v bodě $\boldsymbol{x}$ je potom určena vztahem

$$\hat{Y} = \boldsymbol{\varphi}(\boldsymbol{x})^T \hat{\boldsymbol{w}}_\lambda.$$

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.
- $\varphi(\boldsymbol{x}) = x_i^2$, $\varphi(\boldsymbol{x}) = x_k x_\ell$ – mocniny příznaků a jejich různé součiny, odpovídá polynomiální regresi.

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.
- $\varphi(\boldsymbol{x}) = x_i^2$, $\varphi(\boldsymbol{x}) = x_k x_\ell$ – mocniny příznaků a jejich různé součiny, odpovídá polynomiální regresi.
- $\varphi(\boldsymbol{x}) = \log(x_i), \sqrt{x_i}, \sin(x_i)$ atd. – nelineární transformace jednotlivých příznaků.

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.
- $\varphi(\boldsymbol{x}) = x_i^2$, $\varphi(\boldsymbol{x}) = x_k x_\ell$ – mocniny příznaků a jejich různé součiny, odpovídá polynomiální regresi.
- $\varphi(\boldsymbol{x}) = \log(x_i), \sqrt{x_i}, \sin(x_i)$ atd. – nelineární transformace jednotlivých příznaků.
- $\varphi(\boldsymbol{x}) = \mathbb{1}_{(a,b)}(x_i)$, kde $\mathbb{1}_A(x) = 1$ pokud $x \in A$ a $\mathbb{1}_A(x) = 0$ pokud $x \notin A$ – indikátory množin. Umožňují rozdělení prostoru příznaků na kousky a následné modelování v každém kousku zvlášť.

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.
- $\varphi(\boldsymbol{x}) = x_i^2$, $\varphi(\boldsymbol{x}) = x_k x_\ell$ – mocniny příznaků a jejich různé součiny, odpovídá polynomiální regresi.
- $\varphi(\boldsymbol{x}) = \log(x_i), \sqrt{x_i}, \sin(x_i)$ atd. – nelineární transformace jednotlivých příznaků.
- $\varphi(\boldsymbol{x}) = \mathbb{1}_{(a,b)}(x_i)$, kde $\mathbb{1}_A(x) = 1$ pokud $x \in A$ a $\mathbb{1}_A(x) = 0$ pokud $x \notin A$ – indikátory množin. Umožňují rozdělení prostoru příznaků na kousky a následné modelování v každém kousku zvlášť.
- $\varphi(\boldsymbol{x}) = h(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$, kde $\boldsymbol{x}_i$ je i-tý trénovací bod a h je nějaká funkce – tzv. **radiální bázové funkce** centrované v bodech trénovací množiny.

Bázové funkce

Mezi obvyklé volby bázových funkcí patří:

- $\varphi(\boldsymbol{x}) = x_i$ – přímo jednotlivé příznaky.
- $\varphi(\boldsymbol{x}) = x_i^2$, $\varphi(\boldsymbol{x}) = x_k x_\ell$ – mocniny příznaků a jejich různé součiny, odpovídá polynomiální regresi.
- $\varphi(\boldsymbol{x}) = \log(x_i), \sqrt{x_i}, \sin(x_i)$ atd. – nelineární transformace jednotlivých příznaků.
- $\varphi(\boldsymbol{x}) = \mathbb{1}_{(a,b)}(x_i)$, kde $\mathbb{1}_A(x) = 1$ pokud $x \in A$ a $\mathbb{1}_A(x) = 0$ pokud $x \notin A$ – indikátory množin. Umožňují rozdělení prostoru příznaků na kousky a následné modelování v každém kousku zvlášť.
- $\varphi(\boldsymbol{x}) = h(\|\boldsymbol{x} - \boldsymbol{x}_i\|)$, kde $\boldsymbol{x}_i$ je i-tý trénovací bod a h je nějaká funkce – tzv. **radiální bázové funkce** centrované v bodech trénovací množiny.

Pokud nemáme žádné speciální znalosti o systému, který modelujeme, typicky na počátku volíme velké množství bázových funkcí a používáme hřebenovou regresi, případně jinou formu regularizace.