Lineárne najmenšie štvorce(binárny klasifikátor)

Maximilián Zivák(0.5), Dávid Daniš(0.5)

April 2023

1 Matematické úpravy

1.1 Transformácia sumy do maticovej formy

V zadaní je úloha o najmenších štvorcoch zadaná nasledovne

$$\text{Min} \sum_{i=1}^{N} \left(y^i - \tilde{f}\left(x^i\right) \right)^2 \tag{1}$$

Pričom $\tilde{f}\left(x^i\right) = \left(x^i\right)^T \beta + v$, pre jednoduchosť bude lepšie prepísať samotnú sumu do maticovej formy ktorá vráti $z@1^T$, v ktorom $z^i = \left(y^i - \tilde{f}\left(x^i\right)\right)$. Prvky vo vektore z budú skaláre, keďže y^i je skalar aj $\left(x^i\right)^T \beta$ je skalár. Teda ich môžeme sčítať tak že z vynásobime transponovaným jednotkovým vektorom. Ak zadefinujeme $z = y - z'$ potom vieme zapísať $z' = \tilde{f}$ nasledovne

$$X\beta + v$$

Pričom

$$X = \begin{bmatrix} x_1^1 & x_1^2 & x_1^3 & x_1^4 \\ x_2^1 & x_2^2 & x_2^3 & x_2^4 \\ . & . & . & . \\ x_N^1 & x_N^2 & x_N^3 & x_N^4 \end{bmatrix}$$

Výsledkom násobenie matice X s vektorom β bude vektor ktorého prvky sú riešenia jednotlivých iterácií $\tilde{f}\left(x^i\right) = \left(x^i\right)^T \beta + v$, do ktorého pripočítame skalár v čím dostaneme z'. Pre potreby projektu bude ale lepšie ak v jednom vektore budú obidve premenné (β, v). Ak zadefinujeme nový vektor β' ako:

$$\beta' = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ v \end{bmatrix}$$

Musíme upraviť aj maticu X, keďže výsledok $X\beta + v$ je tvaru

$$\begin{bmatrix} \beta_1 x_1^1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \beta_4 x_1^4 + v \\ . \\ . \\ \beta_1 x_N^1 + \beta_2 x_N^2 + \beta_3 x_N^3 + \beta_4 x_N^4 + v \end{bmatrix}$$

Stačí do matice X pridať stĺpec jednotiek čím dostaneme maticu

$$X' = \begin{bmatrix} x_1^1 & x_1^2 & x_1^3 & x_1^4 & 1 \\ x_2^1 & x_2^2 & x_2^3 & x_2^4 & 1 \\ . & . & . & . & \\ x_N^1 & x_N^2 & x_N^3 & x_N^4 & 1 \end{bmatrix}$$

Ak túto maticu teraz vynásobíme β' získame

$$X'\beta' = \begin{bmatrix} \beta_1 x_1^1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \beta_4 x_1^4 + 1 * v \\ . \\ . \\ \beta_1 x_N^1 + \beta_2 x_N^2 + \beta_3 x_N^3 + \beta_4 x_N^4 + 1 * v \end{bmatrix}$$

Teda $z' = X'\beta' = X\beta + v$. Teraz len dosadíme a získame funkciu

$$\text{Min}((y - X'\beta')\, 1^T)^2$$

1.2 Gradient

Po transformácií úlohy teraz potrebujeme získať gradient na aplikáciu cauchyho metódy. Konkrétne hľadáme gradient funkcie

$$f(\beta') = (y - X'\beta')1^T)^2$$

Aplikovaním chain rule získame

$$2 * (y - X'\beta')1^T) * (y - X'\beta')1^T)'$$

Pred derivovaním druhej časti si ale musím dať pozor lebo vektor $\beta' = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ v \end{bmatrix}$

Teda gradient bude 5 prvkový. Druhú časť si vieme rozpísať pre potreby gradientu ako:

$$-(x_1^1 * \beta_1 + x_1^2 * \beta_2 + x_1^3 * \beta_3 + x_1^4 * \beta_4 + v + ... x_N^1 * \beta_1 + x_N^2 * \beta_2 + x_N^3 * \beta_3 + x_N^4 * \beta_4 + v)$$

(y vynecháme keďže to je konštanta a stále by sa zderivovala na nulu) Je celkom jasne vidieť že pri derivovaní prvkom β_n dostaneme

$$-\sum_{i=1}^{N} x_i^n$$

A pri derivovaní podľa v dostaneme $-N$, teda výsledný gradient bude

$$\nabla f(\beta') = 2 * (y - X'\beta')1^T) * \begin{bmatrix} -\sum_{i=1}^{N} x_i^1 \\ -\sum_{i=1}^{N} x_i^2 \\ -\sum_{i=1}^{N} x_i^3 \\ -\sum_{i=1}^{N} x_i^4 \\ -N \end{bmatrix}$$

2 Úloha a