Predobdelava podatkov

Linearna funkcija

Vhodni podatki so v obliki **X (x_1)** in **Y**, napovedujemo pa funkcijo $\mathbf{y = b_0 + b_1x_1}$. Predobdelava podatkov X v tem primeru **ni potrebna**.

Polinomska funkcija

Vhodni podatki so v obliki **X (x_1)** in **Y**, napovedujemo pa funkcijo $\mathbf{y = b_0 + b_1x_1 + b_2x_2 + \dots b_nx_n}$.

Predobdelava podatkov X **je potrebna**. Sami matriki X moramo dodati stolpce za vsako stopnjo do stopnje *n* ter jim moramo izračunati vrednosti.

Primer: predvidevamo, da imamo polinom 4 stopnje.

Vhodni podatki:

x_1
4

X po obdelavi:

x_1^1	x_1^2	x_1^3	x_1^4
4	16	64	256

Funkcija z več neodvisnimi spremenljivkami

Vhodni podatki so v obliki **X ($x_1, x_2, x_3, \dots, x_n$)** in **Y**, napovedujemo pa funkcijo $\mathbf{y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n}$. Predobdelava podatkov X v tem primeru **ni potrebna**.

Linearna regresija (Metoda najmanjših kvadratov)

Najprej izvedete centriranje podatkov, kar pomeni, da za vsak stolpec X ter Y izračunate povprečje ter vsaki vrednosti v matrikah X in Y odštejete povprečje stolpca kateremu pripada.

Pridobljene matrike uporabite v naslednjem izračunu:

$$\mathbf{b} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

Matrika **b** nam predstavlja koeficiente, matrika X razlagalne spremenljivke in y ciljno.

Izvedete matrično množenje transponirane X matrike z originalno ter zatem naredite inverz rezultata. Za računanje inverzne matrike lahko uporabite poljubno knjižnico! Zatem izvedete še preostala množenja matrik iz te enačbe.

Kot rešitev pridobite koeficiente $b_1, b_2, \dots, b_n$, Izračunati pa morate še koeficient $\mathbf{b_0}$. Izračunate ga tako, v enačbo vstavite za vsak vzorec podatkov $x_1, x_2, .., x_n$ vrednosti ter Y in tako pridobite b_0 za vsak vzorec. Le tem vrednostim potem samo poiščete povprečje in tako dobite $\mathbf{b_0}$ enačbe.