1 Zadanie 1

Drużyna informatyków testuje nowy program rozciągania pleców w przerwach od programowania zespołowego, by ograniczyć liczbę dolegliwości. Poniższa tabelka pokazuje codzienną liczbę minut rozciągania się przez uczestników i powiązaną z nimi liczbę dolegliwości przez cały semestr.

Czas rozciagania	w minutach	0	30	10	15	5	25	35	40
Dolegliwosci		4	1	2	2	3	1	0	1

(a) Znajdź prostą regresji, pokazującą zależność ilości dolegliwości od czasu rozciągania.

(b) O ile spada ilość dolegliwości z każdą minutą rozciągania?

(c) Ile minut rozciągania jest potrzebne, by zawodnik uniknął dolegliwości?

Rozwiązanie:
Niech x,y oznaczają odpowiednio wektory zmiennych objaśniających (czas rozciągania) i zmiennych objaśnianych (liczba dolegliwości). Oznaczmy średnie tych wektorów jako $\bar{x} = 20$, $\bar{y} = 1.75$. Policzmy współczynniki szukanej prostej według wzorów:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = -0.08$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3.35$$

Czyli prosta regresji spełnia równanie $\hat{y} = 3.35 - 0.08x$. Zatem możemy powiedzieć, że z każdą minutą rozciągania liczba dolegliwości spada o 0.08. Aby sprawdzić, ile minut jest potrzebne, by zawodnik uniknął dolegliwości wystarczy podstawić $\hat{y} = 0$ i otrzymujemy

$$\begin{aligned} 0 &= 3.35 - 0.08x \\ 0.08x &= 3.35 \\ x &= 41.875 \end{aligned}$$

Czyli potrzeba co najmniej 41.875 minut.

□

2 Zadanie 2

Pokaż, że jeżeli $n \geqslant p$ oraz $rank(X) = p$ to estymator β metodą najmniejszych kwadratów

$$\hat{\beta} = \arg\min_{\beta} \|Y - X\beta\|^2$$

jest postaci

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

Dodatkowo jeżeli $\varepsilon \sim N(0, \sigma^2 Id)$ to pokaż że $\hat{\beta}$ jest estymatorem największej wiarygodności oraz-znajdź estymator największej wiarygodności σ^2.

Rozwiązanie:
Przepiszmy funkcję błędu średniokwadratowego w postaci

$$RSS(\beta) = \|Y - X\beta\|^2 = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2$$

gdzie X_i oznacza kolumnę macierzy X; oraz zróżniczkujmy ją po β

$$\frac{\partial RSS}{\partial \beta} = -2\sum_{i=0}^{n} X_i^T(Y_i - X_i^T\beta) = -2X^T(Y - X\beta)$$

i drugi raz

$$\begin{aligned}\frac{\partial^2 RSS}{\partial\beta\partial\beta^T} &= -2\frac{\partial}{\partial\beta^T}\left(\sum_{i=0}^{n} X_i^T(Y_i - X_i^T\beta)\right) = -2\frac{\partial}{\partial\beta^T}\left(\sum_{i=0}^{n}(Y_i - \beta^T X_i)X_i^T\right) = 2\sum_{i=0}^{n} X_i X_i^T \\ &= 2X^T X\end{aligned}$$

Skoro X jest pełnego rzędu p, to $X^T X$ dodatnio określona, zatem szukane minimum wyznaczamy przyrównując pierwszą pochodną do zera

$$\begin{aligned}-2X^T(Y - X\hat{\beta}) &= 0 \\ X^T X\hat{\beta} &= X^T Y \\ \hat{\beta} &= (X^T X)^{-1}X^T Y\end{aligned}$$

gdzie $X^T X$ odwracalna, bo X pełnego rzędu.
Niech teraz $\varepsilon \sim N(0, \sigma^2 Id)$. Skoro $\varepsilon_1, ..., \varepsilon_n$ są iid, to mają rozkład $N(0, \sigma^2)$, zatem $Y_1, ..., Y_n$ jest próbką o rozkładzie $N(X_i^T\beta, \sigma^2)$. Korzystając ze wzoru na logarytm funkcji wiarogodności dla rozkładu normalnego mamy

$$\ell(\beta, \sigma) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - X_i^T\beta\right)^2 = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}RSS(\beta)$$

Stąd różniczkując po β otrzymujemy

$$\frac{\partial\ell}{\partial\beta} = -\frac{1}{2\sigma^2}\frac{\partial RSS}{\partial\beta} = 0 \iff \frac{\partial RSS}{\partial\beta} = 0 \iff \beta = \hat{\beta}$$

Czyli $\hat{\beta}$ jest estymatorem największej wiarogodności.

Znajdźmy jeszcze estymator największej wiarogodności σ^2

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} RSS(\beta) = 0 \quad \Longleftrightarrow \quad \sigma^2 = \frac{RSS(\beta)}{n}$$

czyli podstawiając $\beta = \hat{\beta}$

$$\hat{\sigma}^2 = \frac{RSS(\hat{\beta})}{n} = \min_{\beta} \left(\frac{RSS(\beta)}{n} \right)$$

□

3 Zadanie 3

Oblicz wartość oczekiwaną i wariancję $\hat{\beta}$. Jaki rozkład ma $\hat{\beta}$, jeśli $\varepsilon \sim N(0, \sigma^2 Id)$?

Rozwiązanie:
Wiemy, że $\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top Y$ oraz $Y = \mathbf{X}\beta + \varepsilon$. Skoro $\mathbb{E}\varepsilon = 0$, to $\mathbb{E}Y = \mathbf{X}\beta$, zatem

$$\mathbb{E}\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbb{E}Y = \beta.$$

Policzmy teraz wariancję, korzystając z tego, że $\operatorname{Var} \varepsilon = \sigma^2 Id$

$$\begin{aligned} \mathrm{VAR}(\hat{\beta}) &= \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^\top\right] \\ &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbb{E}\left[(Y - \mathbf{X}\beta)(Y - \mathbf{X}\beta)^\top\right] \mathbf{X} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \left(\sigma^2 Id\right) \mathbf{X} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \end{aligned}$$

Załóżmy teraz, że $\varepsilon \sim N(0, \sigma^2 Id)$. Wówczas $Y \sim N(\mathbf{X}\beta, \sigma^2 Id)$, zatem $\hat{\beta}$ jest gaussowski, jako liniowy obraz wektora gaussowskiego Y. Na podstawie wcześniej obliczonej wartości oczekiwanej i wariancji mamy $\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$.

□