ירין אליאב

למידה חישובית - מטלה 1

1 ה- X matrix data, לצורך הנסיבה עמודת המינוס שקיבלנו היא

$$X = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

נבדיק ראשית אם X^TX הפיכה:

$$X^TX = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 7 \end{pmatrix}$$

חישוב $\det(X^TX)$:

$$4 \cdot 7 - 1 \cdot 1 = 27 \neq 0$$

ולכן המטריצה הפיכה. ולכן $\theta^* = X^+y = (X^TX)^{-1}X^Ty$ היא הנק' היחידה שמקיימת $\nabla J(\theta) = 0$, ולכן, אכן יש משמעות למציאת ה pseudo-inverse matrix ל X.

$$X^+ = \begin{pmatrix} 4 & 1 \\ 1 & 7 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{pmatrix}$$

נמצא את $\begin{pmatrix} 4 & 1 \\ 1 & 7 \end{pmatrix}^{-1}$:

כיוון שמדובר במטריצה 2x2.

$$\begin{pmatrix} 4 & 1 \\ 1 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} 7 & -1 \\ -1 & 4 \end{pmatrix} \cdot \frac{1}{27} = \begin{pmatrix} \frac{7}{27} & -\frac{1}{27} \\ -\frac{1}{27} & \frac{4}{27} \end{pmatrix}$$

כעת,

$$X^+ = \begin{pmatrix} \frac{7}{27} & -\frac{1}{27} \\ -\frac{1}{27} & \frac{4}{27} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{8}{27} & \frac{8}{27} & \frac{6}{27} & \frac{5}{27} \\ -\frac{5}{27} & -\frac{5}{27} & \frac{3}{27} & \frac{7}{27} \end{pmatrix} =$$

$$= \frac{1}{27}\begin{pmatrix} 8 & 8 & 6 & 5 \\ -5 & -5 & 3 & 7 \end{pmatrix}$$

נשתמש ב $X^{\dagger}$ בכדי לחשב את θ^* שממזערת את הסכום של ה- squares loss

לפי מה שהסברנו קודם,

$$\theta^* = X^{\dagger} y = (X^T X)^{-1} X^T y = \frac{1}{27}\begin{pmatrix} 8 & 8 & 6 & 5 \\ -5 & -5 & 3 & 7 \end{pmatrix} \underset{\text{לפי הנתונים}}{y} =$$

$$= \frac{1}{27}\begin{pmatrix} 8 & 8 & 6 & 5 \\ -5 & -5 & 3 & 7 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix} \underset{\text{מטריצות כפל}}{=} \frac{1}{27}\begin{pmatrix} 27 \\ 27 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

ולכן $\theta^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ היא הפתרון. כלומר כאשר $\theta_0 = 1$! $\theta_1 = 1$.

נחשב את $J(\theta^*)$:

$$J(\theta^*) = \left((1,1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} + 1\right)^2 + \left((1,1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} - 1\right)^2 + \left((1,1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2\right)^2$$

$$+ \left((1,1)\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3\right)^2 = 1 + 1 + 0 + 0 = 2$$

2 נמצא את ה cost בצורה הבאה:

$$J(\theta) = \sum_{i=1}^{4}\left((\theta_0, \theta_1)\begin{pmatrix} 1 \\ x^{(i)} \end{pmatrix} - y^{(i)}\right)^2 = (\theta_1 - \theta_2 + 1)^2 + (\theta_1 - \theta_2 - 1)^2$$

$$+ (\theta_1 + \theta_2 - 2)^2 + (\theta_1 + 2\theta_2 - 3)^2 = \theta_0^2 - 2\theta_0\theta_1 + \theta_1^2 + 2\theta_0 - 2\theta_1 + 1$$

$$+ \theta_0^2 - 2\theta_0\theta_1 + \theta_1^2 - 2\theta_0 + 2\theta_1 + 1 + \theta_0^2 + 2\theta_0\theta_1 + \theta_1^2 - 4\theta_0 - 4\theta_1 + 4$$

$$\theta_0^2+4\theta_0\theta_1+4\theta_1^2-6\theta_0-12\theta_1+9=4\theta_0^2+2\theta_0\theta_1+7\theta_1^2-10\theta_0-16\theta_1+15$$

לכן, $A=4,\ B=2,\ C=7,\ D=-10,\ E=-16,\ F=15$

נראה שאם $\nabla J(\theta)\in\mathbb{R}^2$ (גרדיאנט) ו $\theta\in\mathbb{R}^2$:

$$\nabla J(\theta)=2\left(\sum_{i=1}^{n}\left(\theta^T x^{(i)}-y_i\right)x_0^{(i)},\ \sum_{i=1}^{n}\left(\theta^T x^{(i)}-y_i\right)x_1^{(i)}\right)$$

$$=2\left(\sum_{i=1}^{n}\left(\theta^T x^{(i)}-y_i\right)x_0^{(i)},\ \sum_{i=1}^{n}\left(\theta^T x^{(i)}-y_i\right)x_1^{(i)}\right)$$

$$=2\left(\left((\theta_0,\theta_1)\begin{pmatrix}1\\-1\end{pmatrix}+1\right)\cdot 1+\left((\theta_0,\theta_1)\begin{pmatrix}1\\-1\end{pmatrix}-1\right)\cdot 1+\right.$$

$$+\left((\theta_0,\theta_1)\begin{pmatrix}1\\1\end{pmatrix}-2\right)\cdot 1+\left((\theta_0,\theta_1)\begin{pmatrix}1\\2\end{pmatrix}-3\right)\cdot 1\Big),\ \left((\theta_0,\theta_1)\begin{pmatrix}1\\-1\end{pmatrix}+1\right)(-1)$$

$$\left.+\left((\theta_0,\theta_1)\begin{pmatrix}1\\-1\end{pmatrix}-1\right)\cdot(-1)+\left((\theta_0,\theta_1)\begin{pmatrix}1\\1\end{pmatrix}-2\right)\cdot 1+\left((\theta_0,\theta_1)\begin{pmatrix}1\\2\end{pmatrix}-3\right)\cdot 2\right)\Big)$$

$$2\left(4\theta_0+\theta_1-5,\ -\theta_0+\theta_1-1-\theta_0+\theta_1+1+\theta_0+\theta_1-2+2\theta_0+4\theta_1-6\right)$$
$$=\left(8\theta_0+2\theta_1-10,\ 2\theta_0+14\theta_1-16\right)$$

נמצא $\nabla J(\theta^*)=0$ כך:

$$\nabla J(\theta^*)=2\left(\left((1,1)\begin{pmatrix}1\\-1\end{pmatrix}+1\right)\cdot 1+\left((1,1)\begin{pmatrix}1\\-1\end{pmatrix}-1\right)\cdot 1+\right.$$

$$\left.+\left((1,1)\begin{pmatrix}1\\1\end{pmatrix}-2\right)\cdot 1+\left((1,1)\begin{pmatrix}1\\2\end{pmatrix}-3\right)\cdot 1\right),\ \left((1,1)\begin{pmatrix}1\\-1\end{pmatrix}+1\right)(-1)$$

$$+\left((1,1)\begin{pmatrix}1\\-1\end{pmatrix}-1\right)\cdot(-1)+\left((1,1)\begin{pmatrix}1\\1\end{pmatrix}-2\right)\cdot 1+\left((1,1)\begin{pmatrix}1\\2\end{pmatrix}-3\right)\cdot 2\Bigg)=$$

$$2\left(1+(-1)+0+0\ ,\ -1+1+0+0\right)=2(0,0)=(0,0)$$

3 נשים לב כי מודל הנסיגה הלינארי היא

$$f_\theta(X)=\theta_0+\theta_1 X_1=1+1\cdot 1.5=2.5$$

התגובה המשוערת נמצאת עם θ^* הוא 2.5.

לפי הנוסחה שראינו בשיעור של k השכנים הקרובים:

$$\hat{y}=\frac{1}{\underset{k=2}{2}}(2+3)=\frac{5}{2}=2.5$$

התגובה המשוערת בנקודה זו היא 2.5.