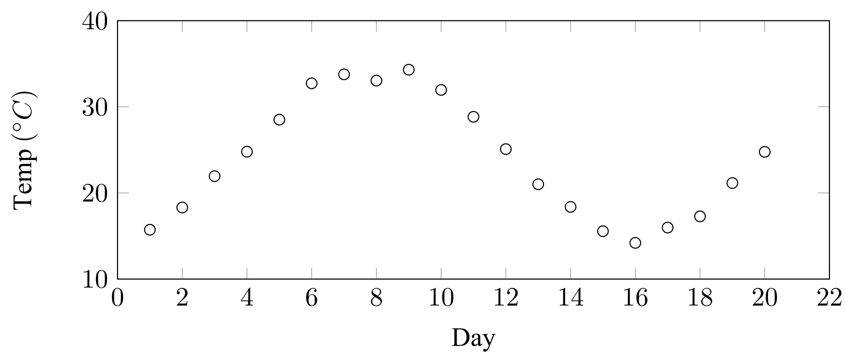
שעור BML 4 - רגרסיה לינארית בייסיאנית

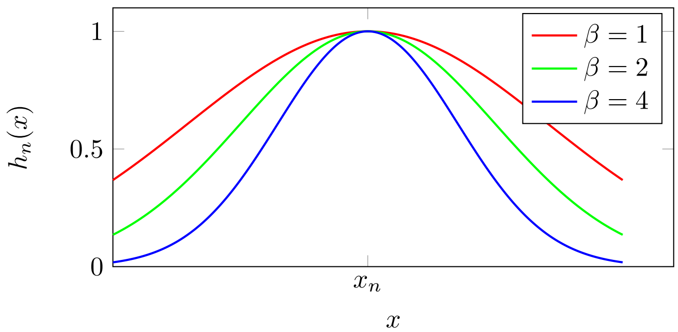
November 17, 2022

הדוגמא שתלווה אותנו הפעם - מדידות מזג אוויר: $D = \{x_i, y_i\}_{i=1}^{I}$ כאשר x_i הוא תאריך (יום 1, יום 2 וכו') ו-y_i הוא הטמפרטורה בירושלים באותו יום.



המטרה שלנו היא לחזות את הטמפרטורה - רוצים לתת תחזית לימים הקרובים, לא ערך יחיד אלא מספר ערכים עם הסתברות לכל אחד.
המודל שלנו הוא $f_\theta(x) = \sum_n \theta_n h_n(x)$ כאשר $\{h_n\}_{n=1}^{N}$ הן פונקציות בסיס ידועות מראש.
דוגמאות לכאלו בסיסים:

1. רגרסיה פולינומיאלית: $h_n(x) = x^n$
2. רשת נוירונים: h_n הוא הנוירון ה-n בשכבה האחרונה של רשת pre-trained ויכול להיות שנרצה קומבינציה לינארית של הייצוגים
3. פונקציות Radial Basis Function - RBF: $h_n(x) = e^{-\beta(x-x_n)}$, כאשר β קובע את רוחב הפעמון



הערה - הפלט $f_\theta(x)$ הוא לינארי ב-θ ומאד לא לינארי ב-x: אם יש לנו $f_{\theta_1}(x), f_{\theta_2}(x)$ אז $f_{(\theta_1+\theta_2)}(x) = f_{\theta_1}(x) + f_{\theta_2}(x)$, לעומת זאת $f_\theta(x_1+x_2) \neq f_\theta(x_1) + f_\theta(x_2)$.

אז איך עושים רגרסיה בייסיאנית? כזכור, המטרה שלנו היא לחשב את $P(\theta|D) = \frac{P(\theta)P(D|\theta)}{P(D)}$ כאשר $P(\theta)$ הוא הפריור, $P(D|\theta)$ היא הנראות ו-$P(D)$ היא הנראות השולית (נקראת גם evidence).

הנחות מודל:

1. יש לנו פריור מהצורה $\theta \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$
2. הנראות היא מהצורה $y(x)|\theta \sim \mathcal{N}(\sum_n \theta_n h_n(x), \sigma^2)$ וכן $y(x_i) \perp y_j(x)|\theta$ (כלומר בהנתן θ "האמיתית" הם בלתי תלויים)

הפריור שלנו במקרים כאלו הוא ההסתברות לקבל מקדמים מסוימים לפונקציה. למשל, ניקח M מדגמים של 200 ימים בהם מזג האוויר בירושלים היה ידוע ולכל אחד מהם נאמוד θ_m, כלומר יש לנו M סטים של מקדמים $\{\theta_m\}_{m=1}^M$ ואז

$$\hat{\mu}_m = \frac{1}{M}\sum_m \theta_m, \qquad \hat{\Sigma}_m = \frac{1}{M}\sum_m (\theta_m - \mu_m)(\theta_m - \mu_m)^T$$

את ההנחה על הנראות ניתן לרשום גם באופן $y(x_i) = \sum_n \theta_n h_n(x) + \eta_i$ כאשר $\eta \sim \mathcal{N}(0, \sigma^2 I)$.
בסך הכל נסמן $y|\theta \sim \mathcal{N}(H\theta, \sigma^2 I)$ כאשר H היא מטריצה בגודל $I \times N$ וכן $H_{i,n} = h_n(x_i)$.

<u>הערות על ההתפלגויות:</u>

1. ההתפלגויות $P(\theta), P(y|\theta)$ גאוסיאניות ולכן גם $P(\theta, y), P(\theta)P(y|\theta)$ גאוסיאניות
2. אם $P(\theta, y)$ גאוסיאנית אז גם $P(\theta|y)$ גאוסיאנית
3. אם $P(\theta, y)$ גאוסיאנית אז $\mu_{\theta|y} = \arg\max_\theta P(\theta, y)$ כי המקסימום של גאוסיאן הוא בתוחלת. הוכחה:

$$\mu_{\theta|y} = \arg\max_\theta P(\theta|y) = \arg\max_\theta \frac{P(\theta, y)}{P(y)} = \arg\max_\theta P(\theta, y)$$

4. המטריצה $Cov(\theta|y)$ הופכית ונגדית לנגזרת השניה של לוג ההתפלגות המשותפת, $\Sigma_{\theta|y}^{-1} = -\frac{\partial^2}{\partial\theta^2}\ln(P(\theta, y))$. הוכחה:

$$P(\theta|y) = \frac{1}{Z_\theta} e^{-\frac{1}{2}(\theta-\mu_{\theta|y})^T \Sigma_{\theta|y}^{-1}(\theta-\mu_{\theta|y})}, \qquad P(y) = ?$$

$$P(\theta, y) = P(\theta|y)P(y) \to \ln\left(P(\theta, y)\right) = \ln\left(P(\theta|y)\right) + \ln\left(P(y)\right)$$
$$= \ln(Z_\theta) - \frac{1}{2}(\theta-\mu_{\theta|y})^T \Sigma_{\theta|y}^{-1}(\theta-\mu_{\theta|y}) + \ln\left(P(y)\right)$$
$$\frac{\partial}{\partial\theta}\ln\left(P(\theta, y)\right) = -\Sigma_{\theta|y}^{-1}(\theta-\mu_{\theta|y})$$
$$\frac{\partial^2}{\partial\theta^2}\ln\left(P(\theta, y)\right) = -\Sigma_{\theta|y}^{-1}$$

משפט : אם $y|\theta \sim \mathcal{N}(H\theta, \sigma^2 I)$ ו-$\theta \sim \mathcal{N}(\mu_\theta, \Sigma_\theta)$ אז $(\theta|y) \sim \mathcal{N}(\mu_{\theta|y}, \Sigma_{\theta|y})$ כאשר

$$\mu_{\theta|y} = \left(\frac{1}{\sigma^2}H^T H + \Sigma_\theta^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}H^T y + \Sigma_\theta^{-1}\mu_\theta\right), \qquad \Sigma_{\theta|y} = \left(\frac{1}{\sigma^2}H^T H + \Sigma_\theta^{-1}\right)^{-1}$$

הוכחה :

$$P(\theta) = \frac{1}{Z_\theta} e^{-\frac{1}{2}(\theta-\mu_\theta)^T \Sigma_\theta^{-1}(\theta-\mu_\theta)}, \qquad P(y|\theta) = \frac{1}{Z_y} e^{-\frac{1}{2}(H\theta-y)^T\left(\sigma^2 I\right)^{-1}(H\theta-y)} = \frac{1}{Z_y} e^{-\frac{1}{2\sigma^2}\|H\theta-y\|^2}$$

כלומר אנחנו מחפשים את

$$\arg\max_\theta \frac{1}{Z_\theta Z_y}\exp\left(-\frac{1}{2}\left((\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta) + \frac{1}{\sigma^2}\|H\theta-y\|^2\right)\right)$$

הערך שממקסם את הפונקציה ממקסם גם את הלוג שלה :

$$= \arg\max_\theta -\frac{1}{2}(\theta-\mu_\theta)^T\Sigma_\theta^{-1}(\theta-\mu_\theta) - \frac{1}{2\sigma^2}\|H\theta-y\|^2 = J(\theta)$$
$$\frac{\partial}{\partial\theta}J(\theta) = -\Sigma_\theta^{-1}(\theta-\mu_\theta) - \frac{1}{\sigma^2}H^T(H\theta-y) = -\left(\frac{1}{\sigma^2}H^T H + \Sigma_\theta^{-1}\right)\theta + \left(\frac{1}{\sigma^2}H^T y + \Sigma_\theta^{-1}\mu_\theta\right)$$
$$\frac{\partial}{\partial\theta}J(\theta) = 0 \to \left(\frac{1}{\sigma^2}H^T H + \Sigma_\theta^{-1}\right)\theta = \left(\frac{1}{\sigma^2}H^T y + \Sigma_\theta^{-1}\mu_\theta\right)$$
$$\to \mu_{\theta|y} = \left(\frac{1}{\sigma^2}H^T H + \Sigma_\theta^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}H^T y + \Sigma_\theta^{-1}\mu_\theta\right)$$

ולפי הערה 4 נמצא את $\Sigma_{\theta|y}$:

$$\frac{\partial}{\partial\theta}\ln\left(P(\theta,y)\right)=-\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)\theta+\left(\frac{1}{\sigma^2}H^Ty+\Sigma_\theta^{-1}\mu_\theta\right)$$

$$\frac{\partial^2}{\partial\theta^2}\ln\left(P(\theta,y)\right)=\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)\to\Sigma_{\theta|y}=\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)^{-1}$$

נשים לב כי לכל מספר דוגמאות, המטריצה $\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}$ הפיכה : המטריצה Σ_θ היא מטריצת שונויות ולכן PD (וגם ההופכית שלה PD), המטריצה H^TH סימטרית ולכן PSD, מכאן כי הסכום $\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)$ הוא מטריצה PD ולכן הפיכה לכל מספר דוגמאות.
במקרה הזה $\theta^{MAP}=\theta^{MMSE}$, לעומת זאת

$$\theta^{MLE}=\arg\max_\theta P(y|\theta)=\arg\max_\theta\frac{1}{Z_\theta}\exp\left(-\frac{1}{2\sigma^2}\|H\theta-y\|^2\right)=\arg\min_\theta\|H\theta-y\|^2$$

$$\frac{\partial}{\partial\theta}\|H\theta-y\|^2=2H^T(H\theta-y)$$

$$\frac{\partial}{\partial\theta}\|H\theta-y\|^2=0\to H^TH\theta=H^Ty\to\theta^{MLE}=\left(H^TH\right)^{-1}H^Ty$$

זה דורש הפיכות של H^TH, כלומר כן נדרש לנו מספר דגימות מינימלי. מתי θ^{MMSE} מתלכד עם θ^{MLE}? נניח כי $\sigma^2\to 0$ וכן H^TH הפיכה, נקבל $\mu_{\theta|y}\to\left(H^TH\right)^{-1}H^Ty$.