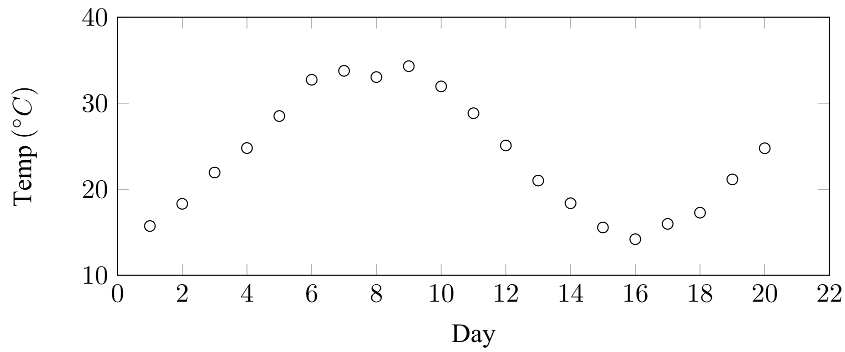
שעור 6 BML - רגרסיה עם ריבוי פונקציות בסיס

November 29, 2022

יש לנו אוסף נקודות $D=\{x_i,y_i\}$ - x_i הוא התאריך ו-$y(x_i)$ הטמפרטורה ביום הזה.



כרגיל $y(x)=\sum_n \theta_n h_n(x)$, אם יש לנו פריור $\theta\sim\mathcal{N}(0,\Sigma_\theta)$ אז $\theta^{MMSE}=\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)^{-1}\frac{1}{\sigma^2}H^Ty$.
יש לנו N פונקציות בסיס ו-I דוגמאות $D=\{x_i,y_i\}_{i=1}^{I}$, המטריצה $C_{\theta|y}=\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)^{-1}$ היא בגודל $N\times N$, כלומר כדי לחשב את θ^{MMSE} עבור בסיס פולינום ממעלה 9 צריך להפוך מטריצה בגודל 10×10.
בגישה הקלאסית נדרוש $I >> N$: למשל עבור בסיס P_9 יש 10 פרמטרים חופשיים לכן נדרוש לפחות 100 דגימות. לעומת זאת, בגישה הבייסיאנית אנחנו מאפשרים גם מצב של $N >> I$.
דוגמא: תמונות ממצלמות רכב - $x\in\mathbb{R}^m$ תמונות של 100^2 פיקסלים, $y\in\mathbb{R}$ מרחק של רכב מהמצלמה בתמונה x.
נגדיר פונקצית בסיס באופן $\left\{h_{mn}(x)\triangleq x(m)x(n)\right\}$, זה סימטרי ולכן יהיו לנו $N=\frac{n(n+1)}{2}=\binom{10^4}{2}\cdot 10^4\approx\frac{10^4\cdot 10^4}{2}=5\cdot 10^7$ כלומר 50 מיליון פונקציות בסיס.
בעיה טיפוסית: יש לנו $I=10^3$ דוגמאות ו-$N=50\cdot 10^6$ פיצ'רים. חישוב של θ^{MMSE} דורש היפוך של מטריצה בגודל $N\times N$ וכך גם החישוב של $C_{\theta|y}$.
הכלי הראשון שיש לנו להתמודד איתו הוא למת היפוך המטריצה (ידועה גם בתור Woodbury Matrix Identity):

$$\theta^{MMSE}=\Sigma_\theta H^T\left(H\Sigma_\theta H^T+\sigma^2 I\right)^{-1}y$$

המכפלה $H\Sigma_\theta H^T$ היא בגודל $I\times I$ וכבר צמצמנו את הקושי החישובי בהיפוך, אבל הוקטור θ עדיין בגודל N והמטריצה $C_{\theta|y}$ עדיין בגודל $N\times N$.
משפט הדואליות: לכל בעית רגרסיה לינארית בייסיאנית פרימאלית עם

$$y_i = y_\theta(x) + \eta, \quad y_\theta(x) = \sum_n \theta_n h_n(x), \quad \theta \sim \mathcal{N}(0, \Sigma_\theta), \quad \eta \sim \mathcal{N}\left(0, \sigma^2\right)$$

קיימת בעית רגרסיה לינארית דואלית עם

$$y_i = y_\alpha(x) + \eta, \quad y_\alpha(x) = \sum_i \alpha_i k(x, x_i), \quad \alpha \sim \mathcal{N}\left(0, K^{-1}\right), \quad \eta \sim \mathcal{N}\left(0, \sigma^2\right)$$

כאשר

$$k(x, x_i) \triangleq h^T(x) \Sigma_\theta h(x_i), \qquad K_{i,j} \triangleq k(x_i, x_j)$$

ומתקיים $f_{\alpha^{MMSE}}(x) = f_{\theta^{MMSE}}(x)$ לכל x.

הוכחה:

יש לנו $y_\alpha(x) = \sum_i \alpha_i k(x, x_i)$, ביחס לנוסחא של $\theta^{MMSE} = C_\theta H^T \left(H C_\theta H^T + \sigma^2 I\right)^{-1} y$ נציב הפעם

$$C_\alpha = K^{-1}, \qquad H = H^T = K, \qquad y = K\alpha + \eta$$

ונקבל:

$$\alpha^{MMSE} = K^{-1} K \left(K K^{-1} K^T + \sigma^2 I\right)^{-1} y = \left(K + \sigma^2 I\right)^{-1} y$$

אנחנו יודעים כי $f_{\alpha^{MMSE}}(x) = \sum_i \alpha_i^{MMSE} k(x, x_i)$ וכן $f_{\theta^{MMSE}}(x) = h^T(x) \theta^{MMSE}$, נציב:

$$\begin{aligned} f_{\theta^{MMSE}}(x) &= h^T(x) \theta^{MMSE} = h^T(x) C_\theta H^T \left(H C_\theta H^T + \sigma^2 I\right)^{-1} y \\ &= h^T(x) C_\theta H^T \left(K + \sigma^2 I\right)^{-1} y = h^T(x) C_\theta H^T \alpha^{MMSE} = \sum_i \alpha_i^{MMSE} k(x, x_i) = f_{\alpha^{MMSE}}(x) \end{aligned}$$

נחזור לדוגמא: $h_{mn}(x) \triangleq x(m)x(n)$ וכן $y_\theta(x) = \sum_{m,n} h_{mn}(x) \theta_{mn}^{MMSE}$, לעומת זאת $y_\alpha(x) = \sum_i \alpha_i k(x, x_i)$. על פניו, כדי לחשב את $k(x, x_i)$ צריך לחשב את $h(x), h(x_i)$ שהם וקטורים באורך $5 \cdot 10^7$. נסמן את x_i בתור דוגמת אימון ו-x דוגמא חדשה, נניח כי $C_\theta = I$ ונקבל:

$$\begin{aligned} k(x, x_i) &= h^T(x) h(x_i) = \sum_{m,n} x(m)x(n)x_i(m)x_i(n) \\ &= \sum_m x(m)x_i(m) \sum_n x(n)x_i(n) = \sum_m x(m)x_i(m)\left(x^T x_i\right) = \left(x^T x_i\right)^2 \end{aligned}$$

וזו כבר עלות ריבועית בגודל התמונה.
הטריק הזה נקרא The Kernel Trick: אם אלגוריתם למידה ניתן לכתיבה רק ע"י חישוב של הפונקציה $k(x, x_i)$ אז אין צורך לחשב את $h(x), h(x_i)$.
דרישות מ-k:

1. סימטריה: $k(x, x_i) = k(x_i, x)$
2. חיוביות: $k(x_i, x_i) > 0$
3. המטריצה K חיובית ממש

לכן למשל $k = \sin^3(x - x_i)$ לא יכול להיות קרנל, כי $k(x_i, x_i) = \sin^3(x_i - x_i) = \sin^3(0) = 0$.