רגרסיה לינארית בייזיאנית - המשך

מה קורה כאשר יש הרבה פונק' בסיס?

<u>דוגמא:</u> $\{X_i\}_{i=1}^{P=1000}$ תמונות של רכב בגודל 100×100 פיקסלים.

y_i המרחק של הרכב מהמצלמה.

נניח שפונק' הבסיס - $h_{m,n}(X_i)=X_i(m)\cdot X_i(n)$ ונגדיר

את $y(X_i)=\sum_{m,n}\theta_{m,n}\cdot h_{mn}(X_i)$. כאשר מניחים פריור

$\theta\sim N(0,\Sigma_\theta)$.

נשים לב שבמקרה זה מספר פונקציות הבסיס הוא 10^8.

כדי לקבל את המשערך הטוב ביותר ל- θ, נשתמש במה

שמצאנו בשיעור הקודם -

$$\mu_{\theta|D}=\theta^{MMSE}=\left(\frac{1}{\sigma^2}H^TH+\Sigma_\theta^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}H^Ty\right)$$

חסר כאן $\Sigma_\theta^{-1}\mu_\theta$ אבל מניחים $\mu_\theta=0$

נשים לב שעלינו להפוך כאן מטריצה בגודל $10^8\times10^8$

בתרגול ראינו ניסוח אלטרנטיבי ל- θ^{MMSE}

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

בצורה הזו, אנחנו צריכים להפוך מטריצה שהיא $10^3 \times 10^3$.
עם זאת, אנחנו עדיין צריכים לעבוד עם Σ_θ שהיא מטריצה
גדולה מאוד. אנחנו רוצים למצוא דרך שקולה שלא
תדרוש מאיתנו לעבוד עם גדלים שתלויים במספר פונק'
הבסיס, אלא רק את מס' הדוגמאות.

<u>משפט:</u> הדואליות של רגרסיה לינארית בייסיאנית

נניח שהבעיה הפרימאלית היא

$$f_\theta(x) = \sum_{n=1}^{N} \theta_n h_n(x) \qquad \theta \sim N(0, \Sigma_\theta)$$

ונגדיר את הבעיה הדואלית

$$f_\alpha(x) = \sum_{i=1}^{p} \alpha_i K(x, x_i) \qquad \alpha \sim N(0, K^{-1})$$

כאשר $K_{p \times p}(i,j) \triangleq h(x_i)^T \Sigma_\theta h(x_j)$

$$K(x, x_i) \triangleq h(x)^T \Sigma_\theta h(x_i)^T$$ -!

אזי מתקיים $f_\theta^{MMSE}(x) = f_\alpha^{MMSE}(x)$

הוכחה-

נזכיר שאם חישבנו את θ^{MMSE}, אז בהינתן דוגמא חדשה x, נחשב את הפרדיקציה עליה בעזרת

$$y_{\theta^{MMSE}}(x) = h^T\theta^{MMSE} = h^T(x)\Sigma_\theta H^T\alpha^*$$

מהניסוח האלטרנטיבי ל- θ^{MMSE}

$$= \sum_i \alpha_i^* h^T(x)\Sigma_\theta h(x_i)$$

$$= \sum_i \alpha_i^* k(x,x_i)$$

כאשר $\alpha^* = (H\Sigma_\theta H^T+\sigma^2 I)^{-1}y = (K+\sigma^2 I)^{-1}y$

הבאנו כאן שאפשר לכתוב את $y_{\theta^{MMSE}}(x) = \sum_i \alpha_i^* k(x,x_i)$

נותר להראות ש- $\alpha^{MMSE} = \alpha^*$.

זו המטריצה K מההגדרה של ה-cov של α (אפשר לוודא).

הבעיה הדואלית בכתיב וקטורי-

$$y_{\theta^{MMSE}} = K\alpha \quad , \quad \alpha \sim N(0,\Sigma_\alpha)$$

אז במקרה הזה* $\alpha^{MMSE} = \left(\frac{1}{\sigma^2}K^TK+\Sigma_\alpha^{-1}\right)^{-1}\frac{1}{\sigma^2}K^Ty$

* K על תקן H בבעיית הרגרסיה, וזה פשוט הפתרון שמצאנו קודם.

ונשים לב ש- $\Sigma_\alpha = K$ ולכן

$$\alpha^{MMSE} = \left(\frac{1}{\sigma^2}K^TK+\Sigma_\alpha^{-1}\right)^{-1}\frac{1}{\sigma^2}K^Ty = \left[\left(\frac{1}{\sigma^2}K^T+I\right)K\right]^{-1}\cdot\frac{1}{\sigma^2}K^Ty$$

$$= \left(\frac{1}{\sigma^2}K^T+I\right)^{-1}\cdot\frac{1}{\sigma^2}y = (K^T+\sigma^2 I)^{-1}y = \alpha^* \quad \blacksquare$$

שמצאנו קודם

נחזור לדוגמא: $h_{mn}(X_i)=X_i(m)\cdot X_i(n)$

$$y(x)=\sum_{m,n}\theta_{m,n}\cdot h_{mn}(x)$$

$$\theta\sim N(0,I)$$

בבעיה הדואלית נחשב $\alpha_{1000\times1}$ ע"י היפוך מטריצה

מגודל 1000×1000 - $\alpha^*=(K+\sigma I)^{-1}y$

אבל אנחנו עדיין צריכים לחשב את המטריצה K,

כאשר כל תא במטריצה דורש מלכתחילה את

$K(i,j)=\sum_{m,n}h_{mn}(x_i)\cdot h_{mn}(x_j)$ - כלומר את $h(x_j),h(x_i)$

נשים לב ש-

$$K(X_i,X_j)=\sum_{mn}h_{mn}(x_i)\cdot \underbrace{I}_{\Sigma_\theta}\, h_{mn}(x_j)$$

$$=\sum_{mn}X_i(m)\cdot X_i(n)\cdot X_j(m)\cdot X_j(n)=(X_i^TX_j)^2$$

לכן ניתן לחשב את המטריצה K בלי לייצר את $h(x)$ לכל דוגמא x.

הגדרה - "The Kernel trick"
מצב שבו ניתן לחשב את $k(x_i,x_j)$ בלי
שמחשבים את $h(x_i), h(x_j)$.

הערה - במצבים שאפשר להשתמש ב- Kernel trick
אפשר לחשב את y_θ^{MMSE} בעלות שהיא
ריבועית במס' הדוגמאות, ולא תלויה כלל במס'
פונקציות הבסיס.