1 תאורטי

תאורטי

תרגיל 1-2.

1. נגדיר את Y להיות המטריצה בה השורה הi ית מתאימה ל y_i. ובאופן דומה נגדיר את H ו ξ ביחס ל וקטורים $h\left(x_i\right)$ נקבל כי

$$\begin{aligned}\log\left(Y|\Theta,\sigma\right) &= \sum_i^N \log\mathcal{N}\left(y_i|\Theta^\top h\left(x_i\right), I\sigma^2\right)\\ &= \sum_i^N \frac{1}{2\sigma^2}\left\langle y_i-\Theta^\top h\left(x_i\right), y_i-\Theta^\top h\left(x_i\right)\right\rangle\\ &= \frac{1}{2\sigma^2}Tr\left(\left(Y-\Theta^\top H\left(x\right)\right)\left(Y-\Theta^\top H\left(x\right)\right)^\top\right)\end{aligned}$$

הביטוי הזה נוח כי $Trace$ היא פעולה לניארית ולכן $\frac{\partial}{\partial\alpha}Tr\left(f\right)=Tr\left(\frac{\partial}{\partial\alpha}f\right)$ כלומר

$$\begin{aligned}\nabla\log\left(Y|\Theta,\sigma\right) &= \frac{1}{\sigma^2}Tr\left(\nabla\left(Y-\Theta^\top H\left(x\right)\right)\left(Y-\Theta^\top H\left(x\right)\right)^\top\right)\\ &= \frac{1}{\sigma^2}Tr\left(\left(Y-\Theta^\top H\left(x\right)\right)H^\top\left(x\right)\right)\\ \frac{1}{\sigma^2}&Tr\left(YH^\top\left(x\right)-\Theta^\top H\left(x\right)H^\top\left(x\right)\right)=0\end{aligned}$$

פתרון אחד אפשרי הוא ש-

$$Tr\left(H\left(x\right)H^\top\left(x\right)\Theta\right)=\sigma^2 Tr\left(Y^\top H\left(x\right)\right)$$

או במילים אחרות ניתן להציג כל איבר על האלכסון של $H\left(x\right)H^\top\left(x\right)\Theta$ כלהיות שווה ל לאיבר על האלכסון של $Y^\top H\left(x\right)$.

2. נוכל כמובן, לשטח את Y כך שיהיה וקטור מהצורה $Y\in\mathbb{R}^{Nk\times 1}$ במקרה שכזה. גם $\Theta\in\mathbb{R}^{pk\times 1}$ ונקבל בעיה רגרסיה פשוטה. כמו במקרה הקודם. נקבל כי $HH^\top$ היא תיהיה מטריצת בלוקים ונוכל להציג את הפתרון הסופי כסכום של k פתרונות.

תרגילים 3-4

1. נשים לב כי העובדה שעבור סט $A=\left\{x_i\right\}^N$ המטריצה $K\left(x_i,x_j\right)$ מתפרקת ל $f_A^\top\left(x_i\right)f_A\left(x_j\right)$ נובעת ישירות מכך שK היא PSD. מאחר ו K סימטרית אז קיימת לה פירוק מהצורה $K=A^\top A$ ולכן נגדיר את $f\left(x_i\right)=A^{(i)}$ ואכן כל לראות ש $K_{ij}=A_i^\top A_j=f\left(x_i\right)^\top f\left(x_j\right)$.
מכן כמובן כל לראות כי

$$\begin{aligned}k_1\left(x,y\right)k_2\left(x,y\right) &= f^\top\left(x\right)f\left(y\right)g^\top\left(x\right)g\left(y\right)=\left(\sum_i f_i\left(x\right)f_i\left(y\right)\right)\left(\sum_i g_i\left(x\right)g_i\left(y\right)\right)\\ &= \sum_i\sum_j f_i\left(x\right)f_i\left(y\right)g_j\left(x\right)g_j\left(y\right)=\left(\sum e_{ij}f_i\left(x\right)g_j\left(x\right)\right)^\top\left(\sum e_{ij}f_i\left(y\right)g_j\left(y\right)\right)\end{aligned}$$

2. נגדיר את $K_\times\left(x,y\right)=\left(\sum e_{ij}f_i\left(x\right)g_j\left(x\right)\right)^\top\left(\sum e_{ij}f_i\left(y\right)g_j\left(y\right)\right)$ ובזה סיימנו.

תרגילים 5-6-7.

1. נסתכל על סט הנקודות בגודל 2 $\left\{x_i\right\}=\left\{x,y\right\}$ (נקבע את x,y בסוף) ונראה כי $K\left(x_i,x_j\right)$ אינה PSD. המטריצה המתקבלת היא:

$$K\left(\left\{x,y\right\}\right)=\begin{bmatrix}e^{(x-x)^2} & e^{(x-y)^2}\\ e^{(y-x)^2} & e^{(y-y)^2}\end{bmatrix}=\begin{bmatrix}1 & \xi\\ \xi & 1\end{bmatrix}$$

נמצא את הע"ע ונקבל כי

$$det\left(K-\lambda I\right)=0\Rightarrow\left(1-\lambda\right)^2=\xi^2\Rightarrow\lambda=1\mp\xi=1\mp e^{(x-y)^2}$$

מאחר ו e^{x^2} אינה חסומה, נוכל לבחור x,y כך של K יהיה ע"ע שלילי בסתירה לכך שהיא PSD.

2. מתכונת כפל בקבוע אנו יודעים כי לכל קרנל k גם $c\cdot k$ הוא קרנל לכל $c>0$ מיצד שני. נניח בשלילה שהטענה נכונה ונסתכל על $k-ck=\left(1-c\right)k$ ל $c>1$ כל שהוא. יהיה $\left\{x_i\right\}$ סט של n נקודות. קל ליראות כי

$$Tr\left(\left(1-c\right)K\left(\left\{x_i\right\}\right)\right)=\left(1-c\right)Tr\left(K\left(\left\{x_i\right\}\right)\right)$$

מאחר ו $K\left(\left\{x_i\right\}\right)$ הוא PSD אז כל הע"ע שלו גדולים או שווים מ 0 ולכן $Tr\left(K\left(\left\{x_i\right\}\right)\right)\geq 0$ נבחר את k המקורי כך שהאישיון יהיה חזק. ונקבל מכאן ש $Tr\left(\left(1-c\right)K\left(\left\{x_i\right\}\right)\right)\leq 0$ כלומר ל $\left(1-c\right)K\left(\left\{x_i\right\}\right)$ יש ערך עצמי שלילי ולכן אינו PSD.

3. מגיע באופן מיידי. סט של $\{\{x_{ia}, x_{ib}\}_i\}^N$ נקודות מגדיר את K_{k_a}, K_{k_b} וברור כי

$$K\left(x, x'\right) = k\left(x, x'\right) = k_a\left(x, x'\right) + k_b\left(x, x'\right) = K_{k_a}\left(x, x'\right) + K_{k_b}\left(x, x'\right)$$

ונקבל כי לכל $z \in \mathbb{R}^N$ מתקיים כי

$$z^\top K z = z^\top \left(K_{k_a} + K_{k_b}\right) z = z^\top K_{k_a} z + z^\top K_{k_b} z \geq 0$$

ולכן K היא גם PSD.

תרגיל 8:

אנו רוצים להראות כי