שעור חזרה – Kernels

גירסת הגרעין של בעיית SVM דואלית

נתבונן בגרסת הגרעין של הבעיה הדואלית ($C=1$):

$$\max_{\alpha\in\mathbb{R}^n}\left(\sum_{k=1}^{n}\alpha_k-\frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}\alpha_k\alpha_l y_k y_l K(x_k,x_l)\right) \quad (1)$$
$$\text{subject to:}\quad 0\le\alpha_k\le 1\quad k=1,\dots,n$$
$$\sum_{k=1}^{n}\alpha_k y_k=0$$

כאשר K הוא גרעין גאוסייני: $K(x,z)=\exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$.

א. הוכח כי הנורמה של כל קלט במרחב המאפיינים $\|\phi(x)\|$ שווה ל-1.

ב. הוכח כי הזווית בין כל שני וקטורי קלט שונים במרחב המאפינים $\phi(x),\phi(y)$ קטנה מ-90.

נגדיר סדרה $S_1=\{(x_k,y_k),k=1,\dots,n\}$, וסדרה $S_2=\{(Ax_k,y_k),k=1,\dots,n\}$.
סדרה S_2 מתקבלת ע"י הכפלת הדגימות של S_1 במטריצה **אורתנורמאלית** A.

א. נניח כי $\hat{\alpha}^1$ הוא הפתרון האופטימאלי של בעיה (1) תוך שימוש בסט האימון S_1, ו- $\hat{\alpha}^2$ הוא הפתרון האופטימאלי של בעיה (1) תוך שימוש בסט האימון S_2. הוכח כי $\hat{\alpha}^1=\hat{\alpha}^2$. האם תכונה זו נכונה לכל גרעין?

ב. נתון כי $h_1(x)$ הוא הסיווג של דוגמא חדשה x ע"י המסווג שאומן ע"י סט האימון S_1, ו- $h_2(x)$ הוא הסיווג של דוגמא זו ע"י מסווג שאומן ע"י סט האימון S_2. לשם סעיף זה בלבד, הנח כי הקבוע b של המסווגים שווה לאפס.
איזו אחת מהטענות הבאות בהכרח נכונה, הוכח טענה זו:

i. $\forall x,\quad h_1(x)=h_2(x)$

ii. $\forall x,\quad h_1(x)=h_2(Ax)$

iii. $\forall x,\quad h_1(Ax)=h_2(x)$

<u>פתרון</u>

א. הוכח כי הנורמה של כל קלט במרחב המאפינים שווה ל-1:

$$\left\|\phi(x)\right\|^2 = \left\langle \phi(x), \phi(x) \right\rangle = K(x,x) = \exp\left(\frac{\left\|x-x\right\|^2}{2\sigma^2}\right) = 1$$

ב. הוכח כי הזווית בין כל שני וקטורי קלט שונים קטנה מ-90:

$$(1) \quad \left\langle \phi(x), \phi(y) \right\rangle = K(x,y) = \exp\left(-\frac{\left\|x-y\right\|^2}{2\sigma^2}\right) > 0$$

$$angle(\phi(x),\phi(y)) = \arccos\left(\frac{\left\langle \phi(x), \phi(y) \right\rangle}{\left\|\phi(x)\right\|\left\|\phi(y)\right\|}\right) = \arccos\left(\frac{\left\langle \phi(x), \phi(y) \right\rangle}{1 \cdot 1}\right) \overset{(1)}{<} 90$$

ג. הוכח כי $\hat{\alpha}^1 = \hat{\alpha}^2$

נשים לב כי בפתרון האופטימאלי תלוי בסט הדגימות רק דרך הביטוי $K(x_i, x_j)$.

לכן מספיק להראות כי: $\forall x, z \qquad K(x,z) = K(Ax, Az)$

$$K(Ax,Az) = \exp\left(-\frac{\left\|Ax-Az\right\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\left(A(x-z)\right)^T \left(A(x-z)\right)}{2\sigma^2}\right)$$

$$= \exp\left(-\frac{\left(x-z\right)^T A^T A\left(x-z\right)}{2\sigma^2}\right) = \exp\left(-\frac{\left(x-z\right)^T I\left(x-z\right)}{2\sigma^2}\right) =$$

$$\exp\left(-\frac{\left(x-z\right)^T \left(x-z\right)}{2\sigma^2}\right) = K(x,z)$$

באופן כללי תכונה זו אינה קיימת לכל גרעין. הסיבה לקיום תכונה זו בגרעין הגאוסיאני נובעת מהאינווריאנטיות של הנורמה ביחס למטריצות אורתונורמאליות.

ד. טענה $h_1(x) = h_2(Ax) \qquad \forall x,$

יש להוכיח שמתקיים

$$h_1(x) = sign\left(w_1{}^T \Phi(x)\right) = sign\left(w_2{}^T \Phi(Ax)\right) = h_2(Ax)$$

$$w_2{}^T \Phi(Ax) = \sum_{i=1}^{m} \alpha_i y_i \Phi(Ax_i) \Phi(Ax) = \sum_{i=1}^{m} \alpha_i y_i K(Ax_i, Ax) \underset{(1)}{=} \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) =$$

$$= \sum_{i=1}^{m} \alpha_i y_i \Phi(x_i) \Phi(x) = w_1{}^T \Phi(x)$$