גישות חישוביות – תרגיל 4

.1

.a

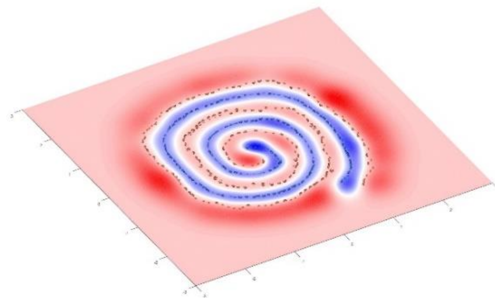
λ הוא המקדם לסכום על ξ כדי להעניש דוגמאות שסווגו לא נכון. באופן כללי, מייצר איזון בין שתי אפשרויות:

λ גדול לא מאפשר דוגמאות סוררות, ה$margin$ יהיה קטן.

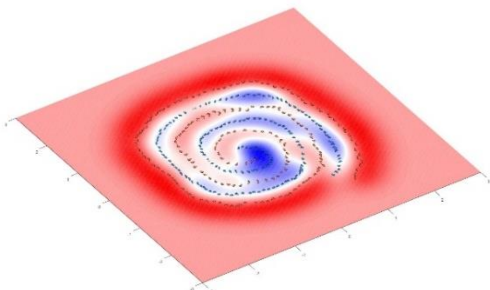
λ קטן לא נענש על דוגמאות סוררות, ה$margin$ יכול להיות גדול יותר.

מהרצת ה$data$ של אופציה 5 (סט של שתי ספירלות), שאינו ניתן להפרדה ליניארית, ראינו שλ קטן מאפשר טעויות על הדוגמאות וλ גדול מייצר הפרדה קשיחה.

λ גדול:



λ קטן:



.b

<u>קלט:</u>

H – מטריצת קרנל סימטרית בגודל $n*n$ כאשר n גודל ה$data\ set$, מקיימת:

$$H(i,j)=H(j,i)=y_i\ *y_j*K(x_i,x_j)=y_i\ *y_j*e^{-\beta\left|\left|x_i-x_j\right|\right|^2}$$

f – תנאי ליניארי, בתרגיל שלנו וקטור מינוס אחדות בגודל n.

lb – חסם תחתון על וקטור המשתנים z שמחפשים כדי לפתור את המשוואה, כלומר $lb\le z$, בתרגיל $0\le z$.

ub – חסם עליון על וקטור המשתנים z שמחפשים כדי לפתור את המשוואה, כלומר $z\le ub$, בתרגיל $z\le\infty$.

Aeq – וקטור התיוגים של הדוגמאות, בגודל $n*1$.

Beq – ערך המכפלה הנדרש עבור $Aeq*z$ (כלומר צריך להתקיים $Aeq*z=Beq$ עבור ה z הנבחר).

<u>פלט:</u>

מחזירה את וקטור המקדמים z בגודל $n*1$.

<u>אופן הפעולה:</u>

הפונקציה מאתרת את וקטור המקדמים z שפותר את המשוואה : $\min_z\ \frac{1}{2}*z^THz+f^Tz$, תחת האילוצים $lb\le z\le ub,\ Aeq*z=Beq$.

.c

ה $bias$ מחושב לפי הנקודות שנבחרו כ $near\ boundery$. עבור וקטור nb של לכל היותר n נקודות (כאשר סט האימון מכיל n דוגמאות) הפונקציה מחשבת:

1. לכל i ב nb :
$$\frac{1}{y(i)} - f(x) \rightarrow temp(i)$$
2. $bias = mean(temp)$.

כלומר עבור תיוגים $\{-1,1\}$ בבעיית הכרעה בינארית הפונקציה מחשבת את ה$bias$ להיות הממוצע של ההפרשים בין התיוג האמיתי לחיזוי עבור הדוגמאות של nb.

.d

$SVMtrial(x, y, kw, \lambda)$:

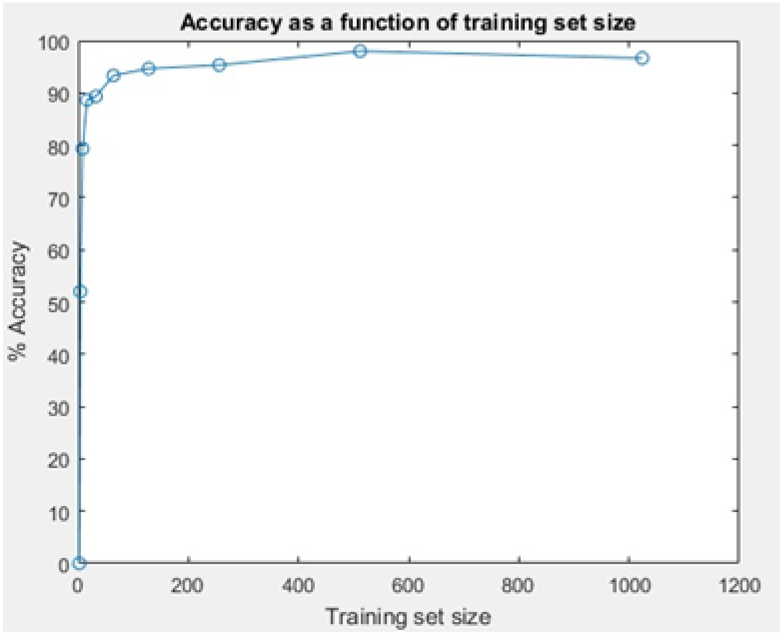
1. נרמל את x.
2. $x \rightarrow xT$.
3. $\begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} \rightarrow f\ ,\ Aeq \rightarrow y^T\ , Beq \rightarrow 0\ ,\ lb \rightarrow (0, \dots, 0)\ , ub \rightarrow \lambda$
4. הפק מטריצת קרנל H
   a. לכל j , $1 \le j \le n$:
      i. לכל i , $1 \le i \le j$:
         1. $x(j) - xT(i) \rightarrow d$
         2. $y_j y_i K(x_i, x_j) \rightarrow H(j,i)$ // בתרגיל $K(x_i, x_j) = e^{-\frac{||d||^2}{kw}}$
         3. $H(j,i) \rightarrow H(i,j)$
5. הרץ $quadratic\ program\ (H, f, Aeq, Beq, lb, ub) \rightarrow\ \ a$
6. עדכן $support\ vectors\ (sv)$ ע"מ לאפשר טעות קטנה ($0 < a < \varepsilon\ \ or\ \lambda < a < \lambda - \varepsilon$)
7. חשב את ה$bias(b)$ (כפי שתואר בסעיף 1.c).
8. החזר F כאשר המכיל את x, y, sv, kw, a, b .

.2

a. מומש.

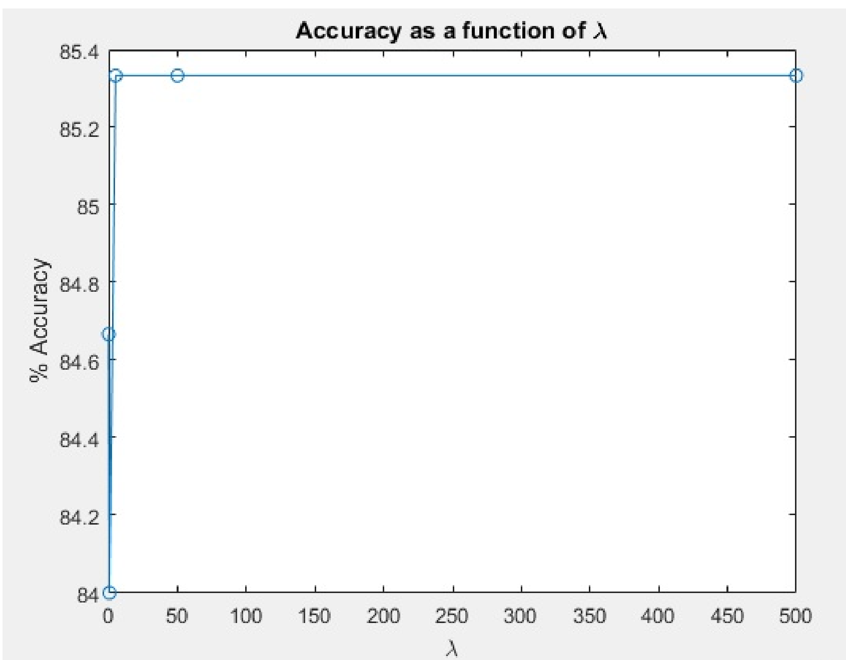
b. עבור 3 החלוקות של ה$data\ set$ קיבלנו אחוזי דיוק של 86,84,92 , אחוז הדיוק הממוצע $87\frac{1}{3}$.

c. ניתן לראות שככל שמגדילים את כמות הדוגמאות ב $training\ set$ אחוזי הדיוק על סט הולידציה עולים. בנוסף, עבור $training\ set$ בגודל 2 ה $bias$ חושב כNaN מכיוון שלא נמצאו $support\ vectors$, דבר שפגע ביצירת התיוג, ולכן אחוז הדיוק הוא 0.



d. קיבענו $training\ set$ בגודל 10 דוגמאות וסט ולידציה של 50 דוגמאות, להלן אחוזי הדיוק כפונקציה של λ :

λ	0.005	0.5	5	50	500
$Accuracy(\%)$	84.667	84	85.33	85.33	85.33



ניתן לראות בתוצאות שמתקבלים אחוזי דיוק דומים עבור ערכי הλ השונים. ייתכן שנובע מסט אימון קטן מדי ולכן הענישה עבור דוגמאות סוררות לא באה לידי ביטוי.

.3

נניח $data\ set$ של n דוגמאות מתויגות $(x_1, y_1), \dots, (x_n, y_n)$. תהי מטריצה H בגודל $n * n$,
$H = \begin{bmatrix} y_1y_1K(x_1,x_1) & \cdots & y_1y_nK(x_1,x_n) \\ \vdots & \ddots & \vdots \\ y_ny_1K(x_n,x_1) & \cdots & y_ny_nK(x_n,x_n) \end{bmatrix}$ המייצגת את מכפלת תיוגי שתי דוגמאות רצופות בתוצאת
ה$kernel$ על הדוגמאות. נייצר וקטור משתנים $Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ ותנאי ליניארי $f = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}$ (נבחר בתרגיל).

כעת נראה שקילות בין פיתרון ה $quadprog$ לפיתרון SVM בעזרת $slack\ variables$.
נעשה זאת על ידי הוכחת שקילות בין $quadprog$ לבעיה הדואלית יחד עם שקילות בין הבעיה הדואלית לבעיה הפרימלית.

א. שקילות בין פיתרון ה$quadprog$ לבעיה הדואלית:

$$\min_z \ \frac{1}{2} * z^T H z + f^T z \qquad , \ such\ that\ \ lb \le z \le ub\ , \ \ Aeq * z = Beq, \qquad where\ Aeq = y^T$$

$$\Leftrightarrow$$

$$\min_z \ \frac{1}{2} * (z_1, \dots, z_n) \begin{bmatrix} y_1y_1K(x_1,x_1) & \cdots & y_1y_nK(x_1,x_n) \\ \vdots & \ddots & \vdots \\ y_ny_1K(x_n,x_1) & \cdots & y_ny_nK(x_n,x_n) \end{bmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} + (-1, \dots, -1) \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

$$\Leftrightarrow$$

$$\min_z \ \frac{1}{2} * (z_1, \dots, z_n) \begin{bmatrix} y_1y_1K(x_1,x_1) & \cdots & y_1y_nK(x_1,x_n) \\ \vdots & \ddots & \vdots \\ y_ny_1K(x_n,x_1) & \cdots & y_ny_nK(x_n,x_n) \end{bmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} - \sum_{j=1}^{n} z_j$$

$$\Leftrightarrow$$

$$\min_z \ \frac{1}{2} * (z_1, \dots, z_n) \begin{pmatrix} \sum_{i=1}^{n} y_1 y_i K(x_1, x_i) z_i \\ \vdots \\ \sum_{i=1}^{n} y_n y_i K(x_n, x_i) z_i \end{pmatrix} - \sum_{j=1}^{n} z_j$$

$$\Leftrightarrow$$

$$\min_z \ \frac{1}{2} * \left(\sum_{i=1}^{n} y_1 y_i K(x_1, x_i) z_1 z_i + \cdots + \sum_{i=1}^{n} y_1 y_i K(x_1, x_i) z_n z_i \right) - \sum_{j=1}^{n} z_j$$

$$\Leftrightarrow$$

$$\min_z \ \frac{1}{2} * \left(\sum_{j=1}^{n} \sum_{i=1}^{n} y_j y_i K\left(x_j, x_i\right) z_j z_i \right) - \sum_{j=1}^{n} z_j$$

מקסימום על X שקול
למינימום על X-.

$$\Leftrightarrow$$

$$\max_z \ -\left(\frac{1}{2} * \left(\sum_{j=1}^{n} \sum_{i=1}^{n} y_j y_i K\left(x_j, x_i\right) z_j z_i \right) - \sum_{j=1}^{n} z_j \right)$$

$$\Leftrightarrow$$

$$\max_z \ \sum_{j=1}^{n} z_j - \frac{1}{2} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} y_j y_i K\left(x_j, x_i\right) z_j z_i \right) \qquad , \quad such\ that\ :\ \sum_{j=1}^{n} z_j y_j = 0$$

ב. כעת נותר להראות שקילות בין הבעיה הדואלית לבעיה הפרימלית, כלומר :

$$\max_z \sum_{j=1}^{n} z_j - \frac{1}{2}\left(\sum_{j=1}^{n}\sum_{i=1}^{n} y_j y_i K(x_j, x_i) z_j z_i\right) \Leftrightarrow \min \frac{1}{2}\left||w|\right|^2 + C\sum_{i=1}^{n}\xi_i$$

כאשר עבור הבעיה הפרימלית: $\sum_{i=1}^{n}\xi_i \geq 0$, $(w^T x_i + b)y_i \geq 1 - \xi_i$,
רוצים להביא למינימום את w , $w = \sum_{j=1}^{n} z_j x_j y_j$.

עבור הבעיה הדואלית דורשים: $\sum_{j=1}^{n} z_j y_j = 0$

הוכחה:

$$\min \frac{1}{2}\left||w|\right|^2 + C\sum_{i=1}^{n}\xi_i \ , \qquad such\ that: (w^T x_i + b)y_i \geq 1 - \xi_i , \sum_{i=1}^{n}\xi_i \geq 0 \ , \quad where\ \ w = \sum_{j=1}^{n} z_j x_j y_j$$

$\Leftrightarrow$

$$\max_{\mathrm{z}} \qquad \min_{\mathrm{w,b}} \qquad \frac{1}{2}\left||w|\right|^2 - \sum_{i=1}^{n} z_i((w^T x_i + b)y_i - 1)$$

$\Leftrightarrow$

$$\max \frac{1}{2}\sum_{i=1}^{n} z_i x_i y_i \sum_{j=1}^{n} z_j x_j y_j - \sum_{i=1}^{n} z_i\left(\left(\sum_{j=1}^{n} z_j x_j y_j\, x_i + b\right) y_i - 1\right)$$

$\Leftrightarrow$

$$\max \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i y_i z_j x_j y_j + \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} z_i\left(\left(\sum_{j=1}^{n} z_j x_j y_j\, x_i + b\right) y_i\right)$$

$\Leftrightarrow$

$$\max \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i y_i z_j x_j y_j + \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} b z_i y_i - \sum_{i=1}^{n} z_i y_i x_i \sum_{j=1}^{n} z_j x_j y_j$$

$\Leftrightarrow$

$$\max \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i y_i z_j x_j y_j + \sum_{i=1}^{n} z_i - \sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i y_i z_j x_j y_j$$

$\Leftrightarrow$

$$\max \sum_{i=1}^{n} z_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i y_i z_j x_j y_j \qquad such\ that: \sum_{j=1}^{n} z_j y_j = 0 \qquad \blacksquare$$

.4

בחרנו להחליף את פונקציית ה$kernel$ שהייתה בקוד, RBF, בפונקציית הקרנל הגאוסיאנית:

$$K(x_1, x_2) = -\frac{||x_1 - x_2||^2}{2\sigma^2}$$

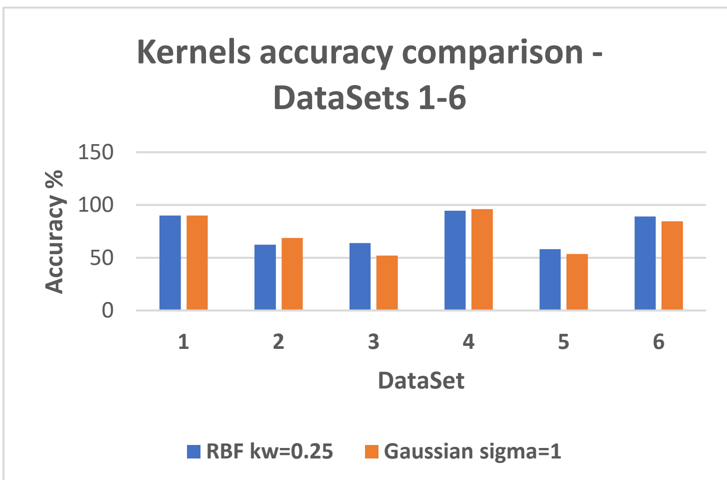
מימשנו את החלפת הפונקציה והרצנו על גבי הסטים 1-6 הנתונים. לטובת ההיפרפרמטר σ בחרנו ערכים שונים בניתוחים השונים.

ביצענו שני ניתוחים:

א. השוואת אחוזי הדיוק עבור $Data\ Sets$ שונים בין שני הקרנלים (RBF עם $kw = \frac{1}{4}$ ו - $Gaussian$ עם $\sigma = 1$).
ראינו שאחוזי הדיוק נשארים יחסית דומים ללא מגמה.

ב. השוואת אחוזי הדיוק עבור $Data\ Set$ רנדומלי עם תיוגים שנקבעים לפי חוקיות (4) עבור σ משתנה. ראינו שערכי σ של 0.5,1 נותנים את הביצועים הטובים ביותר, כשהגדלה והקטנה פוגעות בהדרגה באחוזי הדיוק.

להלן התוצאות:



Data Set	Name
1	Linearly Seperable
2	Non-Linearly Seperable
3	Random Set
4	Random Set - Labels Determined By Rule
5	Two Spirals
6	Imbalance