שאלה 1

תזכורת:

$$f(w_1,\dots,w_k)=\frac{1}{n}\sum_{i=1}^{n}\max_{j\in[k]}\left(x_i(w_j-w_{y_i})+\mathbb{1}(j\neq y_i)\right)$$

[illegible] מינימום (של f) [illegible] כך [illegible] $\hat{y}_i=\hat{j}$ [illegible]

[illegible] f [illegible]. נניח בשלילה [illegible] אם f [illegible]

[illegible] $w_1^*,\dots,w_k^*$ (כך ש) $w^*=(w_1^*,\dots,w_k^*)$ [illegible] [illegible]

אזי, לכל $j\in[k]\setminus\{y_i\}$ כך ש:

$$w_j\cdot x_i\ge w_{y_i}x_i \Rightarrow x_i(w_j-w_{y_i})\ge 0 \Rightarrow x_i(w_j-w_{y_i})+1>0$$

ולכן f לא [illegible] [illegible].

[illegible], [illegible] f, [illegible] f

[illegible] [illegible].

נסמן ב- $w_1^*,\dots,w_k^*$:

$$y_i=\arg\max_j w_j^* x_i \Rightarrow \forall j: w_{y_i}^* x_i \ge w_j^* x_i \Rightarrow x_i(w_j^*-w_{y_i}^*)\le 0$$

ונסמן:
$$z_i=\min_{j\in[k]}(w_{y_i}^*-w_j^*)x_i$$
$$z=\min_{i\in[n]} z_i$$

[illegible]

$$\forall j\in[k]\setminus\{y_i\}: (w_j-w_{y_i})x_i=\frac{1}{z}(w_j^*-w_{y_i}^*)x_i=\frac{-1}{z}(w_{y_i}^*-w_j^*)x_i \le \underset{\text{מהגדרת } z_i}{\uparrow}\frac{-z_i}{z}\le -1$$

[illegible] [illegible] [illegible] [illegible] f [illegible] [illegible].

□

שאלה 2

נזכור הבעיה של SVM הדואלית, ללא הגבלות לבעיה הפרימלית:

$$\max_{\lambda=(\lambda_1,\dots,\lambda_n)} \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j y_i y_j x_i x_j \quad \text{s.t } \lambda_i \geq 0\ \forall i,\ \sum_{i=1}^{n}\lambda_i y_i = 0$$

בבעיה שלנו יחידה כי $n=2$, $y_1=1\ y_2=-1$ והבעיה $\lambda=(\lambda_1,\lambda_2)$ ונקבל:

$$\max_{\lambda=(\lambda_1,\lambda_2)} \lambda_1+\lambda_2 - \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\lambda_i\lambda_j y_i y_j x_i x_j$$

ובנוסף לפי האילוץ: $\lambda_1 y_1+\lambda_2 y_2=0$ ולכן נקבל ש: $\underline{\lambda_1=\lambda_2}$ ונקבל:

$$\max_{\lambda_1} 2\lambda_1 - \frac{1}{2}\left(\lambda_1^2\left(\|x_1\|^2+\|x_2\|^2\right)\right) + \lambda_1^2\langle x_1,x_2\rangle$$

ומשווים ל-0

נגזור לפי λ_1 מה סימטרי

$$2-\lambda_1^*\left(\|x_1\|^2+\|x_2\|^2\right)+2\lambda_1^*\langle x_1,x_2\rangle=0$$

$$2=\lambda_1^*\|x_1-x_2\|^2$$

$$\lambda_1^*=\frac{2}{\|x_1-x_2\|^2}=\lambda_2^*$$

מאחר סימטרי

ועכשיו לפי המשוואות בקשר / דואליות נוכל למצוא את w,b:

$$w=\sum_{i=1}^{2}\lambda_i^* x_i y_i=\lambda_1 x_1\cdot 1+\lambda_1 x_2\cdot(-1)=\lambda_1(x_1-x_2)=\boxed{\frac{2(x_1-x_2)}{\|x_1-x_2\|^2}}$$

$$\lambda_1^*\left(1-y_1(wx_1+b)\right)=0$$: complementary slackness

אבל $\lambda_1^*\neq 0$ כי $x_1\neq x_2$ ולכן לפי משוואה זו נקבל את השוויון

$$1-y_1(wx_1+b)=0$$

$$b=y_1-wx_1=\boxed{1-\frac{2x_1(x_1-x_2)}{\|x_1-x_2\|^2}}$$

וקיבלנו את w,b האופטימליים ■

שאלה 3

ⓐ נכון. [illegible] ξ_i [illegible] C [illegible] $0 \leq \xi_i$ [illegible]

[illegible]

[illegible]

ⓑ [illegible] נגדיר [illegible]

[illegible] $0 < (\xi_i)^2$ [illegible]

אם [illegible] ξ_i [illegible] $\sum_{i=1}^{n} \xi_i$ [illegible]

[illegible] : האילוץ [illegible] ξ_i

[illegible] $y_i(w^T x_i + b) \geq 1 - \xi_i$ [illegible] C [illegible]

אם [illegible] ξ_i [illegible] 0, האילוץ [illegible]

[illegible]

[illegible] האילוצים $0 \leq \xi_i$ מיותרים.

□

ⓑ ~~[illegible]~~

~~$\nabla f = (w_1, \dots, w_n, \dots)$~~

ⓑ ~~$\nabla f = (w_1, 0, \dots, C, \dots, C)$~~

~~$\nabla g = (y_i x_i, y_i, \dots, 1)$~~

~~$L(w, b, \xi, \alpha) = f(w, b, \xi) + \sum \alpha_i (1 - \xi_i - y_i(w^T x_i + b))$~~

ⓑ לפי הגדרת הלגרנז'יאן:

$$L(w,b,\xi,\lambda) = f(w,b,\xi) + \sum_{i=1}^{n} \lambda_i \left(1-\xi_i - y_i(w^T x_i + b)\right)$$

~~$= \frac{1}{2}\|w\|^2 - \lambda\cdot\xi + \frac{C}{2}\|\xi\|^2 + \sum_{i=1}^{n} \lambda_i\left(1-y_i(w^T x_i+b)\right)$~~

ⓒ נגזור ונשווה ל-0 לקבלת מינימום:

$$\nabla_w L = w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^{n} \lambda_i y_i x_i$$

$$\nabla_b L = -\sum_{i=1}^{n} \lambda_i y_i = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i y_i = 0 \Rightarrow \langle \lambda, y\rangle = 0$$

$$\nabla_{\xi_i} L = C\xi_i - \lambda_i = 0 \Rightarrow \xi_i = \frac{\lambda_i}{C}$$

נסמן גם ווקטור $\lambda=(\lambda_1,\dots,\lambda_n)$ כך ש $\xi = \frac{\lambda}{C}$.

$$\sum_{i=1}^{n} \lambda_i\left(1-\xi_i - y_i(w^T x_i+b)\right) = \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{n}\lambda_i\xi_i - w^T\sum_{i=1}^{n}\lambda_i y_i x_i - b\underbrace{\sum_{i=1}^{n}\lambda_i y_i}_{=0}$$

[illegible] w [illegible]

ולכן נקבל לפי הגדרת f:

$$\frac{1}{2}\|\underbrace{\sum_{i=1}^{n}\lambda_i y_i x_i}_{\|w\|^2}\|^2 + \frac{\|\lambda\|^2}{2\cdot C} - \|\lambda\|^2 + \sum_{i=1}^{n}\lambda_i = \underbrace{\frac{\|\lambda\|^2}{2C} - \|\lambda\|^2 + \sum_{i=1}^{n}\lambda_i - \frac{1}{2}\|\sum_{i=1}^{n}\lambda_i y_i x_i\|^2}_{\circledast}$$

לכן הבעיה הדואלית המינימום כפונקציה של λ הוא ⊛

□

נציב את הבעיה המקורית בבעיה של הקודם:

$$\min_{\lambda_i} \frac{1}{2}\left\|\sum_{i=1}^{n}\lambda_i y_i x_i\right\|^2 \|w\|^2\left(1-\frac{1}{4}\right) - \sum_{i=1}^{n}\lambda_i$$

ובהינתן ש:

$$\sum_{i=1}^{n}\lambda_i y_i = 0$$

ערכים אופטימליים לפי הבעיה של הדואלית של [illegible].

סעיף 4

יהי w^* פתרון של hardSVM ויהי (v,ξ) פתרון של soft SVM. מתקיים לכל $C \ge \|w^*\|^2$:

$$\frac{1}{2}\|w'\|^2 + \|w^*\|^2\sum_{i=1}^{n}\xi_i \le \frac{1}{2}\|w'\|^2 + C\cdot\sum_{i=1}^{n}\xi_i \le \frac{1}{2}\|w^*\|^2$$

כי w^* פתרון אופטימלי, ל- [illegible] SVM [illegible] soft SVM

ולכן לאחר העברת אגפים:

$$\frac{1}{2}\|w'\|^2 \le \|w^*\|^2\left(\frac{1}{2} - \sum_{i=1}^{n}\xi_i\right)$$

ולכן $\sum_{i=1}^{n}\xi_i \le \frac{1}{2} < 1$ כי אחרת צד ימין יהיה שלילי ונקבל סתירה.

ולכן, לכל i מתקיים $\xi_i \le \frac{1}{2} < 1$. לכן:

$$y_i w_i x_i \ge 1-\xi_i \ge \frac{1}{2} > 0$$

□

שאלה 5

ניעזר במה שראינו שזה גרעין:

$\Phi = \begin{pmatrix} \varphi(x_1)^T \\ \vdots \\ \varphi(x_m)^T \end{pmatrix}$ מטריצה מגודל m, על $d(m)$

ואז מהבינום, נקבל [illegible] לכל [illegible] 0. מתקיים:

$$k_{ij}: K(x_i,x_j) = \sum_{k=0}^{q}\binom{q}{k}(x_i x_j)^k = \sum_{k=0}^{q}\sqrt{\binom{q}{k}}\,x_i^k\cdot\sqrt{\binom{q}{k}}\,x_j^k$$

ולכן נקבל מטריצת ונדרמונדה מהצורה:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^q \\ 1 & x_2 & x_2^2 & \cdots & x_2^q \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^q \end{pmatrix}$$

שדרגתה היא m לפי הנחה, ולכן כל עמודה נכפיל לפי מה שלמדנו בלינארית 1, הדרגה נשמרת, ולכן לכל עמודה j ($j\in\{1,\dots,q+1\}$), נכפיל אותה ב- $\sqrt{\binom{q}{j-1}}$

ולכן נקבל:

$$\Phi = \begin{pmatrix} \sqrt{\binom{q}{0}}\cdot 1 & \sqrt{\binom{q}{1}}\,x_1 & \cdots & \sqrt{\binom{q}{q}}\,x_1^q \\ \vdots & \vdots & & \vdots \\ \sqrt{\binom{q}{0}}\cdot 1 & \sqrt{\binom{q}{1}}\,x_m & \cdots & \sqrt{\binom{q}{q}}\,x_m^q \end{pmatrix}$$

ודרגתה נשארת m

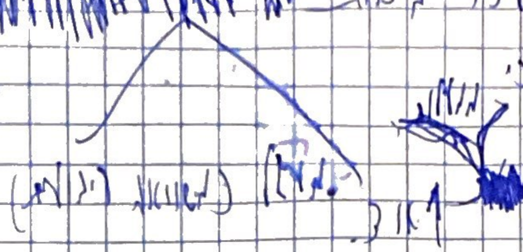
ולכן כמו שראינו בהרצאה נקבל $K=\Phi\cdot\Phi^T$

ולכן הדרגה של K היא m, ואז מהעבודה נקבל ש- לבעיית הארד SVM יש פתרון אופטימלי.

□

שאלה ניתוח

ⓐ ניתן לומר שהמודל עם הטעות מדרגה 2 מודל כי לוקח את הבעיה. הבעיה לינ
היא שהבעיה היא בעיה [illegible] ולכן הטעות גבוה.
2 מודל טוב את הבעיה כי
אך בשביל לייצג מודל [illegible] [illegible]



ⓑ כאן ניתן לומר שהמודל מדרגה 3 נכון המודל יותר טוב את ה-dataset, והוא עדיין
(כמעט) הרעיון הזה טוב יותר. מכיוון שיש שיפור בין הדרגות ניתן
למצוא לאחר האופטימיזציה לנקודה ס (מינימום) ולכן הטעות [illegible]
ולכן המודל [illegible] [illegible] [illegible], טוב יותר. [illegible] [illegible] [illegible]
[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible].

~~ⓒ ניתן לומר שלמרות ה-dataset הקטן, המודל עם RBF הוא הטוב ביותר~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]~~
~~[illegible] [illegible] [illegible] [illegible] [illegible].~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] overfitting~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]~~
~~[illegible] [illegible] [illegible] [illegible] [illegible] [illegible].~~

ⓒ – מודל הגבוה

ⓒ [illegible] ל-[illegible] [illegible]

[illegible]

ל- RBF [illegible] של [illegible]

[illegible] ב- RBF [illegible]

[illegible] של הרעש, [illegible]

אם [illegible]

[illegible] RBF.

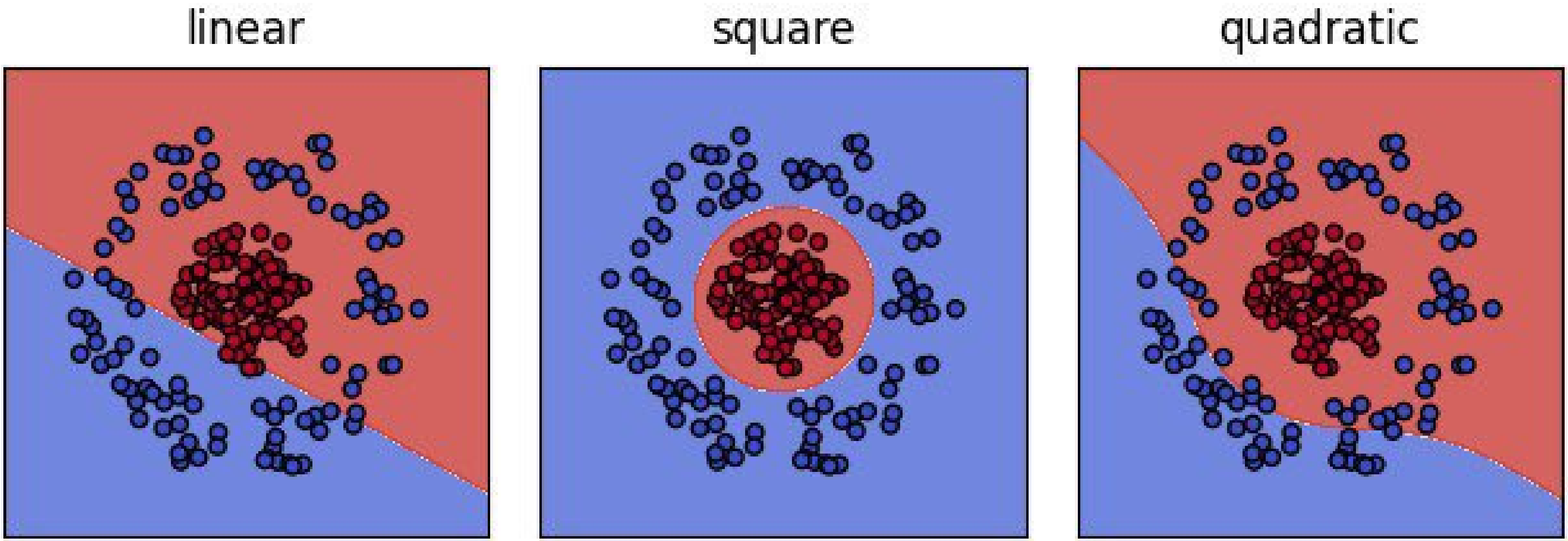
[illegible]

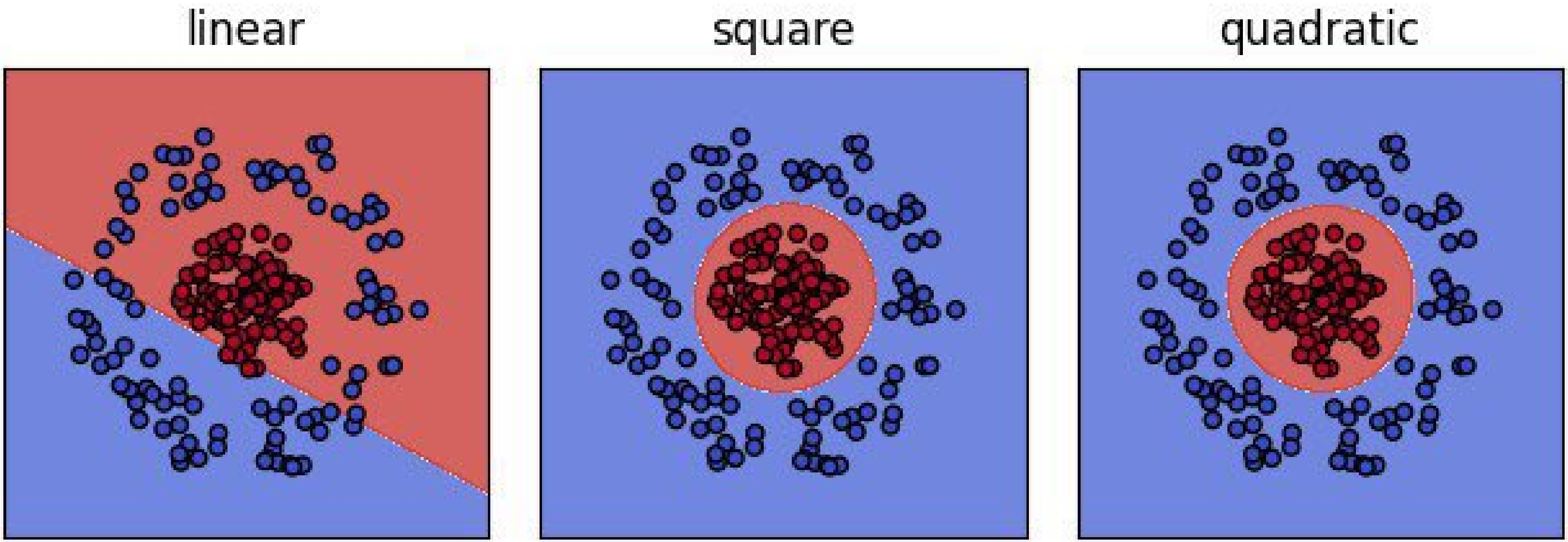
[illegible]

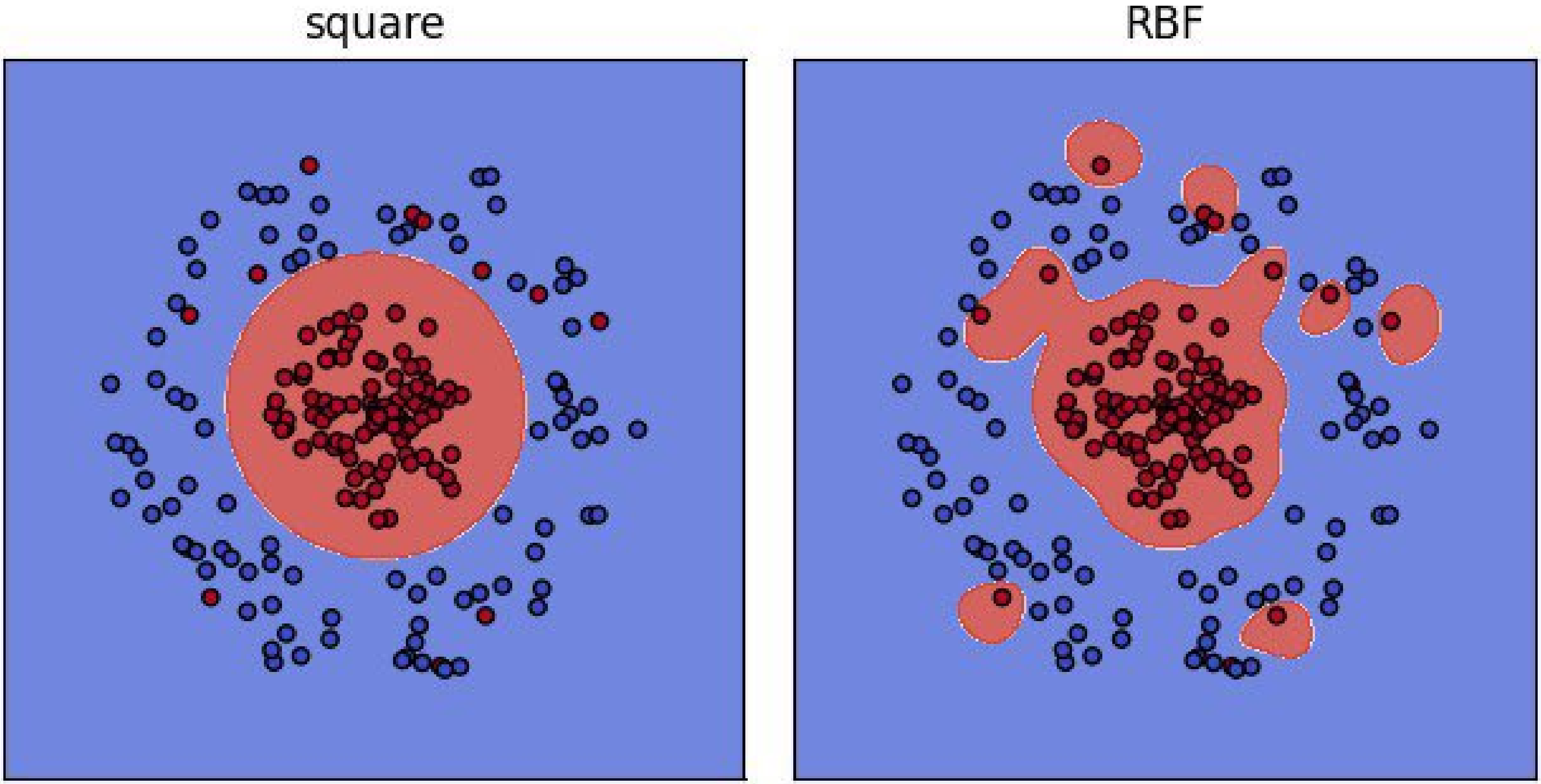
[illegible] overfitting ל-data, [illegible]

[illegible]

[illegible] של [illegible].


linear
square
quadratic


linear
square
quadratic


square
RBF