Machine Learning from Data – IDC – 2022

HW5 – Theory + SVM

1. Kernels and mapping functions (25 pts)
    a. (20 pts) Let $K(x, y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

       Find ψ for which $K$is a kernel. (It may help to first expand the above term on the right-hand side).

    b. (2 pts) What did we call the function ψ in class if we remove all coefficients?
    c. (3 pts) How many multiplication operations do we save by using $K(x, y)$ versus $\psi(x) \cdot \psi(y)$?

2. Lagrange multipliers (25 pts)

   Let $f(x, y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x, y) = \frac{x^2}{4} + y^2 = 1$.

3. PAC Learning (25 pts)

   Let $X = \mathbb{R}^2$. Let vectors $u = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), w = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right), v = (0, -1)$

   and $C = H = \left\{ h(r) = \left\{ (x_1, x_2) \middle| \begin{matrix} (x, y) \cdot u \le r, \\ (x, y) \cdot v \le r, \\ (x, y) \cdot w \le r \end{matrix} \right\} \right\}$, for $r > 0$,

   the set of all origin-centered upright equilateral triangles.
   Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

4. (15 pts) A business manager at your ecommerce company asked you to make a model to predict whether a user is going to proceed to checkout or abandon their cart. You created the model using, and reported 20% error on your test set of size 1000 samples. In the business manager's presentation to upper management, he presented your

model and stated that the company can expect 20% error when deploying the model live on the website.

Luckily, you realize that this is a mistaken assumption, and you correct the statement to say that with 95% confidence, the true error they can expect is up to what percentage? (Just state the error percentage).

5. SVM (10 pts)

   See the notebook in the homework files and follow the instructions there.

   Take a **screenshot** of your resulting graph near the bottom of the notebook (titled “My Graph”) and paste into your submission PDF along with your answers to the theoretical questions. Do **NOT** submit your code.

1. Kernels and mapping functions (25 pts)
   a. (20 pts) Let $K(x, y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

      Find ψ for which K is a kernel. (It may help to first expand the above term on the right-hand side).

a)

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$K(x,y) = (x \cdot y + 1)^3 = (xy)^3 + \binom{3}{2}(x \cdot y)^2 + 3 \cdot x \cdot y + 1 = (xy)^3 + 3(xy)^2 + 3(xy) + 1 =$$

$$(xy)^3 + 3(xy)^2 + 3(xy) + 1$$

$$\left((x_1, x_2)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right)^3 + 3\left(\left((x_1, x_2)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right)^2 \right) + 3(x_1y_1 + x_2y_2) + 1 =$$

$$(x_1y_1 + x_2y_2)^3 + 3\left((x_1y_1 + x_2y_2)^2 \right) + 3(x_1y_1 + x_2y_2) + 1 =$$

$$(x_1y_1)^3 + 3(x_1y_1)^2 x_2y_2 + 3x_1y_1(x_2y_2)^2 + (x_2y_2)^3 + 3\left((x_1y_1)^2 + 2x_1y_1x_2y_2 + (x_2y_2)^2 \right) + 3x_1y_1 + 3x_2y_2 + 1$$

$$(x_1y_1)^3 + 3(x_1y_1)^2 x_2y_2 + 3x_1y_1(x_2y_2)^2 + (x_2y_2)^3 + 3x_1^2y_1^2 + 6x_1y_1x_2y_2 + 3(x_2y_2)^2 + 3x_1y_1 + 3x_2y_2 + 1 =$$

$$x_1^3y_1^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + x_2^3y_2^3 + 3x_1^2y_1^2 + 6x_1y_1x_2y_2 + 3x_2^2y_2^2 + 3x_1y_1 + 3x_2y_2 + 1 =$$

1 2 3 4 5 6 7 8 9 10

$$x_1^3y_1^3 + x_2^3y_2^3 + 3x_1^2y_1^2x_2y_2 + 3x_1y_1x_2^2y_2^2 + 3x_1^2y_1^2 + 3x_2^2y_2^2 + 3x_1y_1 + 3x_2y_2 + 6x_1y_1x_2y_2 + 1 =$$

1 2 3 4 5 6 7 8 9 10

$$\psi(x) = \left(x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, \sqrt{3}x_1^2, \sqrt{3}x_2^2, \sqrt{3}x_1, \sqrt{3}x_2, \sqrt{6}x_1x_2, 1 \right)$$

$$\psi(y) = \left(y_1^3, y_2^3, \sqrt{3}y_1^2y_2, \sqrt{3}y_1y_2^2, \sqrt{3}y_1^2, \sqrt{3}y_2^2, \sqrt{3}y_1, \sqrt{3}y_2, \sqrt{6}y_1y_2, 1 \right)$$

b. (2 pts) What did we call the function ψ in class if we remove all coefficients?

c. (3 pts) How many multiplication operations do we save by using $K(x, y)$ versus $\psi(x) \cdot \psi(y)$?

b) The Rational Veraity

c) $K(x,y) = 4$ multiplications
$\psi(x) \cdot \psi(y) = 46$ multiplications
we saved 42 multiplications

2. <u>Lagrange multipliers (25 pts)</u>

Let $f(x, y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x, y) = \frac{x^2}{4} + y^2 = 1$.

$$g(x,y) = \frac{x^2}{4} + y^2 = 1 \longrightarrow g(x,y) = \frac{x^2}{4} + y^2 - 1$$

$$L(x,y) = 2x - y + \lambda\left(\frac{x^2}{4} + y^2 - 1\right) = 2x - y + \frac{\lambda}{4}x^2 + \lambda y^2 - \lambda$$

$$\frac{\partial}{\partial x} L(x,y) = 2 + \frac{\lambda}{2}x = 0 \qquad (\mathrm{I}$$

פסוק שקר   אם $x=0$ אזי $2=0$ (*)

$$\frac{\partial}{\partial y} L(x,y) = -1 + 2\lambda y = 0 \qquad (\mathrm{II}$$

פסוק שקר   אם $y=0$ אזי $-1=0$ (**)

$$\frac{\partial}{\partial \lambda} L(x,y) = \frac{x^2}{4} + y^2 - 1 = 0 \qquad (\mathrm{III}$$

$$2 + \frac{\lambda}{2}x = 0 \quad /2 \qquad\qquad -1 + 2\lambda y = 0$$

הערה

ש (אם) ! (מצא) ניתן

להציב ב x,y

$$4+\lambda x=0 \qquad 2\lambda y=1$$

$$\lambda x=-4 \qquad \lambda=\frac{1}{2y}$$

$$\lambda=\frac{-4}{x}$$

$$\frac{-4}{x}=\frac{1}{2y}$$

נציב ב III $\boxed{-8y=x}$

$$\frac{(-8y)^2}{4}+y^2-1=0$$

$$16y^2+y^2-1=0$$

$$17y^2=1 \longrightarrow y^2=\frac{1}{17} \longrightarrow y=\pm\sqrt{\frac{1}{17}}$$

עבור $y=\sqrt{\frac{1}{17}}$ נקבל: $x=-8\cdot\sqrt{\frac{1}{17}}=-1.94$

עבור $y=-\sqrt{\frac{1}{17}}$ נקבל: $x=-8\cdot-\sqrt{\frac{1}{17}}=1.94$

ולכן $(x,y)=\left(\sqrt{\frac{1}{17}},-1.94\right) / \left(-\sqrt{\frac{1}{17}},1.94\right)$

נציב אותן ב-f ונקבל:

$$f\left(\sqrt{\frac{1}{17}},-1.94\right)=0.941$$

$$f\left(-\sqrt{\frac{1}{17}},1.94\right)=-2.245$$

כלומר $\left(\sqrt{\frac{1}{17}},-1.94\right)$ היא נק' מקסימום.

ו- $\left(-\sqrt{\frac{1}{17}},1.94\right)$ היא נק' מינימום.

3. <u>PAC Learning (25 pts)</u>

Let $X = \mathbb{R}^2$. Let vectors $u = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), w = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right), v = (0, -1)$

and $C = H = \left\{ h(r) = \left\{ (x_1, x_2) \middle| \begin{array}{l} (x, y) \cdot u \le r, \\ (x, y) \cdot v \le r, \\ (x, y) \cdot w \le r \end{array} \right\} \right\}$, for $r > 0$,

the set of all origin-centered upright equilateral triangles.

Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

נתייחס ל- r כמרחק ממרכז משולש לקודקודיו (שווה כי שווה צלעות)
ולכן עבור כל נקודה ב- $\mathbb{R}^2$ קיים משולש יחידי בו הנקודה נמצאת
על צלע כלשהי שלו. (נשים לב כי r קובע את "גודל" המשולש).

<u>תיאור האלגוריתם:</u>

נעבור על כל הדגימות עבור כל דגימה x כך ש $label(x) = +$, נשמור את
ערך הרדיוס של המשולש שעל אחת מצלעותיו נמצאת הנקודה.
מבין כל הרדיוסים המתקבלים עבור דגימות אלו, ניקח את הרדיוס המקסימלי ונסמן $\bar{r}$.
נשים לב כי המשולש שרדיוסו מתקבל מ- $\bar{r}$ מכיל את כלל הדגימות שה- label שלהן
הוא "+".

נסמן ב- r^* את הרדיוס שמתאר את הקונספט c ($c \in C$ - כאשר C הוא מחלקת הקונספטים),
מכיוון ואין טעויות ב data $\bar{r} \le r^*$ כלומר המשולש שמתאר על r^* גדול או
שווה למשולש המתקבל ע"י $\bar{r}$.
נשים לב שזהו אלגוריתם קונסיסטנטי מאחר וכל הנקודות x שעבורן $label(x)=$"+" נמצאות
הכלולות המשולש (השטח) הזה והוא המשולש עם הרדיוס המינימלי שמקיים את התנאי
והוא מוכל במשולש המתקבל ע"י r^*.

<u>סיבוכיות זמן:</u>

1. מעבר על כלל הדגימות ב- data ומציאת כל הדגימות x שעבורן $label(x)=$"+" , $O(n)$.
2. עבור כל הרדיוסים המתקבלים מ- 1, נמצא את המקסימלי מביניהם ונסמנו $\bar{r}$, $O(n)$.

לכן מ- 1,2 נקבל $O(n) + O(n) = O(n)$.

Sample Complexity:

נרצה למצוא את כמות הדגימות עבורה הסיכוי שנקבל טעות הקטנה מ-ϵ יהיה לפחות δ.

נסמן ב- r^ϵ את הרדיוס המקסימלי כך שההסתברות לדגימה להיות בשטח שבין המשולש המתואר ע"י r^* (הקונספט) לבין המשולש המתואר ע"י r^ϵ תהיה קטנה או שווה ϵ.

נחלק למקרים:

אם $r^\epsilon \le \bar{r}$: אזי סיכוי לדגימה להמצא בין שטח המשולש המתקבל ע"י $\bar{r}$ (ההיפותזה) לבין r^* (הקונספט) ההפרש קטן מ-ϵ.

אם $r^\epsilon > \bar{r}$: אזי הסיכוי של דגימה להיות בשטח שבין המשולש המתואר ע"י $\bar{r}$ (ההיפותזה) לבין r^* (הקונספט) גדול מ-ϵ, כלומר ההסתברות לטעות בגודל מ-ϵ היא $(1-\epsilon)^m$!:

$$m \ge \frac{\ln\left(\frac{1}{\delta}\right)}{\epsilon} \quad \text{כאשר} \quad (1-\epsilon)^m \le \exp(-\epsilon m)$$

ונקבל ש:

$$\exp(-\epsilon m) \le \exp\left(-\ln\left(\frac{1}{\delta}\right)\right) = \exp(\ln(\delta)) = \delta$$

כלומר Sample Complexity פולינומיאלי.

4. (15 pts) A business manager at your ecommerce company asked you to make a model to predict whether a user is going to proceed to checkout or abandon their cart. You created the model using, and reported 20% error on your test set of size 1000 samples. In the business manager's presentation to upper management, he presented your model and stated that the company can expect 20% error when deploying the model live on the website.
Luckily, you realize that this is a mistaken assumption, and you correct the statement to say that with 95% confidence, the true error they can expect is up to what percentage? (Just state the error percentage).

עד 22.5%.

טיוטא

$n=1000$, $r=200$, $se = \sqrt{\frac{0.2(0.8)}{1000}} = 0.012$

$\hat{p} = \frac{200}{1000} = 20\% = 0.2 \qquad (\hat{p}-2se, \hat{p}+2se) = (0.174, 0.225)$

5. SVM (10 pts)

See the notebook in the homework files and follow the instructions there.

Take a **screenshot** of your resulting graph near the bottom of the notebook (titled "My Graph") and paste into your submission PDF along with your answers to the theoretical questions. Do **NOT** submit your code.

Cs = [0.0005, 0.001, 0.0018, 0.00435, 0.01, 0.05, 1]