למידה חישובית 1 (096411)

חורף תשפ"ה 2025

תרגיל בית 4

תאריך אחרון להגשה: 16/01/2025 בשעה 23:59

הוראות הגשה

- ההגשה בזוגות בלבד, דרך "קבוצה" ייעודית שיצרתם במודל.
- עליכם להגיש קובץ **pdf** בודד:
  - `HW4_ID1_ID2.pdf` – קובץ המכיל תשובות לכל השאלות. עבור שאלה 3, יש להוסיף צילומי מסך של הקוד והפלטים שהוא מפיק. ניתן גם לייצא מחברת בפורמט PDF ולשרשר אותה לתשובות לחלק היבש.
- קוד חייב להיות קריא, תמציתי ומתועד היטב. יש להקפיד על שימוש בשמות משמעותיים למשתנים.
- כל גרף חייב להכיל לפחות את האלמנטים הבאים: כותרת, מקרא (legend), כותרות לצירים ויחידות (ticks).
- יש להשתמש בפורום במודל לטובת שאלות על התרגיל. השאלות שלכם עוזרות לסטודנטים אחרים בקורס.

שאלה 1

בהרצאות למדנו על **Regularized Loss Minimization (RLM**) וראינו כיצד באמצעות שיטה זו ניתן לקבל לומדים יציבים ולמתן את תופעת ה-**overfitting**.

תהי $l(w,x,y)$ פונקציה קמורה ב-w (כאשר אנו מתייחסים ל x,y כקבועים). יהי $S=\{(x_i,y_i)\}_i^m$ מדגם אימון ותהי (x',y') תצפית נוספת. בהינתן S, נגדיר את: $f_S(w)=L_S(w)+\lambda\parallel w\parallel^2$ כאשר: $L_S(w)=\frac{1}{m}\sum_{i=1}^m \quad l(w,x_i,y_i)$.

כמו כן נגדיר את $A(S)=argmin_w f_S(w)$. בנוסף, בהינתן $i\in\{1,\dots,m\}$ נגדיר את $S^{(i)}$ באופן הבא:

$$S^{(i)}=\{(x_1,y_1),\dots,(x_{i-1},y_{i-1}),(x',y'),(x_{i+1},y_{i+1}),\dots(x_m,y_m)\}$$

ונגדיר את $A(S^{(i)})=argmin_w f_{S^{(i)}}(w)$.

א. הסבירו במילותיכם מה הוא $S^{(i)}$ ומה הוא $A(S)$.

ב. הסבירו את נכונות השוויון הבא לכל u,v,i:

$$f_S(v)-f_S(u)=L_{S^{(i)}}(v)+\lambda\parallel v\parallel^2-\left(L_{S^{(i)}}(u)+\lambda\parallel u\parallel^2\right)+\frac{l(v,x_i,y_i)-l(u,x_i,y_i)}{m}+\frac{l(u,x',y')-l(v,x',y')}{m}$$

ג. בשימוש הטענה הנ"ל, הסבירו מדוע אי השוויון הבא נכון:

$$f_S\left(A(S^{(i)})\right)-f_S(A(S))\le\frac{l(A(S^{(i)}),x_i,y_i)-l(A(S),x_i,y_i)}{m}+\frac{l(A(S),x',y')-l(A(S^{(i)}),x',y')}{m}$$

ד. הוכיחו כי:

$$\lambda\parallel A(S^{(i)})-A(S)\parallel^2\ \le\frac{l(A(S^{(i)}),x_i,y_i)-l(A(S),x_i,y_i)}{m}+\frac{l(A(S),x',y')-l(A(S^{(i)}),x',y')}{m}$$

ה. הוכיחו כי אם $l(\cdot)$ היא פונקציה $\rho-Lipschitz$ אזי:

$$\parallel A(S^{(i)})-A(S)\parallel\le\frac{2\rho}{\lambda m}$$

ו. הוכיחו כי אם $l(\cdot)$ היא פונקציה $\rho-Lipschitz$ אזי $l(A(S^{(i)}),x_i,y_i)-l(A(S),x_i,y_i)\le\frac{2\rho^2}{\lambda m}$:

ז. כעת נגדיר $L_D(w)=E_{(x,y)\sim D}[l(w,x,y)]$ ו-$L_S(w)=\frac{1}{m}\sum_{i=1}^m \quad [l(w,x_i,y_i)]$. הסבירו במילותיכם מה משמעות כל אחת מההגדרות. קבעו האם בהינתן מדגם אימון S ולומד w ניתן לחשב את ערך הביטוי או לא. נמקו.

ח. כעת הניחו כי $i\sim U(m)$ וכי ניתן לדגום את $(x',y')\sim D$ באופן בלתי תלוי ב-S. הוכיחו כי:

$$E_{S\sim D^m}[L_D(A(S))-L_S(A(S))]\le\frac{2\rho^2}{\lambda m}$$

שאלה 2

בשאלה זו נדון במשמעות של מטריקות שונות (כגון precision) ונדגים שרטוט של עקומת ROC בבעיות סיווג בינאריות.

א. כתבו את הנוסחאות של כל אחת מהמטריקות הבאות: TPR, recall, precision. השתמשו בסימונים מתוך מטריצת הבלבול **(**confusion matrix**)** – TP, TN, FP, FN. הסבירו במילותיכם מה כל מטריקה מייצגת. עבור כל מטריקה, ציינו האם אנו רוצים למקסם או למזער אותה.

ב. תנו דוגמה (במילים) למשימת סיווג בינארית בה ה-recall חשוב יותר מה-precision. הצדיקו את ההצעה שלכם.

ג. תנו דוגמה (במילים) למשימת סיווג בינארית בה ה-precision חשוב יותר מה-recall. הצדיקו את ההצעה שלכם.

ד. כעת הניחו שאימנתם מודל רגרסיה לוגיסטית על מדגם אימון בעל 2 פיצ'רים - x_1, x_2. לאחר האימון התקבלו המשקולות הבאות:

$$w_0 = -0.3, w_1 = -0.5, w_2 = 0.5$$

מדגם האימון נראה כך:

	x1	x2	y
0	8	2	0
1	2	4	1
2	16	2	1
3	4	2	0
4	9	2	0

ד. כתבו את הנוסחה עבור $P_w(y = 1|x)$ עם המשקולות הנ"ל וחשבו את $P_w(y_i = 1|x_i)$ לכל אחת מהתצפיות $i \in \{1, \dots, 5\}$.

ה. שרטטו (בדף ועט או באמצעי אלקטרוני שאיננו תכנותי) את עקומת ה- ROCעבור המודל והתצפיות הנ"ל. יש לשרטט בנוסף את העקומה עבור מודל אקראי. שימו לב כי יש לחשב את ערכי ה-FPR וה-TPR עבור שרטוט זה. פרטו את חישוביכם.

ו. חשבו את ערך המטריקה AUC-ROC, ופרטו את חישוביכם. האם נדמה שהמודל טוב יותר ממודל אקראי? נמקו.

שאלה 3

בשאלה זו ניישם שימוש ב cross validation על סט הנתונים Fashion-MNIST. סט נתונים זה, מכיל כ-70,000 תמונות שחור לבן של 10 סוגים של פרטי לבוש. כל תמונה מלווה בתיוג של סוג פריט הלבוש שמופיע בה.

דוגמה מתוך סט הנתונים (ללא תיוגים):

כל תמונה מיוצגת ע"י מטריצה דו מימדית מגודל 28*x*28 עם ערכים בין 0 ל-255 (כאשר 0 מייצג פיקסל שחור לחלוטין ו-255 מייצג פיקסל לבן לחלוטין). בשאלה זו נעבוד עם ייצוג שטוח של המטריצות הדו מימדיות, כלומר כל תמונה תיוצג ע"י וקטור חד מימדי מגודל 28*28=784. נשתמש בקומבינציות שונות של kernels והיפר-פרמטרים שונים עבור אלגוריתם SVM על מנת לקבוע איזה מסווג צפוי להיות הטוב ביותר.

א. השתמשו בקטע הקוד הבא בכדי לטעון 7000 תמונות ותוויות מתוך סט הנתונים:

```
import numpy as np
from sklearn.datasets import fetch_openml

def fetch_mnist():
    #Download MNIST dataset
    X, y = fetch_openml('Fashion-MNIST', version=1, return_X_y=True)
    X = X.to_numpy()
    y = y.to_numpy()

    # Randomly sample 7000 images
    np.random.seed(2)
    indices = np.random.choice(len(X), 7000, replace=False)
    X, y = X[indices], y[indices]
    return X, y

X, y = fetch_mnist()
print(X.shape, y.shape)
```

יש לוודא שקטע הקוד מדפיס את הפלט הבא: `(7000, 784) (7000,)`. ייתכן והטעינה תיקח מספר שניות.

ב. הציגו את 10 התצפיות (תמונות) הראשונות מהמדגם X באמצעות הפונקציה `plt.imshow` עם הארגומנט `cmap="binary"`. שימו לב כי יש לשנות את הצורה (`reshape`) של כל תצפית ב-X בחזרה למימד 28*28 על מנת להציג אותה באמצעות הפונקציה `imshow`. לכל תמונה, הציגו בסמוך אליה את התיוג המתאים שלה מ-y בצורה הבאה: `(class_index, class_name)`.

אתם יכולים להיעזר במילון הבא:

```
idx2class={'0': 'T-shirt/top', '1': 'Trouser', '2': 'Pullover', '3': 'Dress', '4':
'Coat', '5': 'Sandal', '6': 'Shirt', '7': 'Sneaker', '8': 'Bag', '9': 'Ankle'}
```

ג. ממשו פונקציה בשם `SVM_results(X_train, y_train, X_test, y_test)` כאשר:

- $X_{train} \in R^{m_{train}\times 784}$ – מטריצת הנתונים עבור סט האימון (מטיפוס numpy nd-array)
- $y_{train} \in R^{m_{train}}$ – וקטור הלייבלים עבור סט האימון (מטיפוס numpy nd-array)
- $X_{test} \in R^{m_{test}\times 784}$ – מטריצת הנתונים עבור סט המבחן (מטיפוס numpy nd-array)
- $y_{test} \in R^{m_{test}}$ – וקטור הלייבלים עבור סט המבחן (מטיפוס numpy nd-array)

על הפונקציה להשתמש בפונקציה `cross_validation_error(X, y, model, folds)` שמימשתם בתרגיל בית 3 עם `folds=4` בכדי לחשב את שגיאות האימון והולידציה הממוצעת של מסווגי SVM עם היפר-הפרמטרים הבאים:

- קרנל לינארי עם ערך C ברירת מחדל
- קרנל פולינומי עבור ערכי $d \in \{2, 4, 6, 8\}$
- קרנל RBF עבור ערכי $\gamma \in \{0.001, 0.01, 0.1, 1.0, 10\}$

סה"כ 10 מודלים שונים. בנוסף, לכל מודל מהנ"ל, הפונקציה צריכה להתאים את אותו המודל עבור כל מדגם האימון ולחשב את שגיאת המבחן. הפונקציה צריכה להחזיר מילון (dictionary) כאשר המפתחות (keys) הם שמות המודל (לדוגמא: `'SVM_poly_4'`) והערכים (values) הינם tuple מהצורה הבאה:

`(average_train_error, average_validation_error, test_error)`

כאשר 2 האלמנטים הראשונים מחושבים ע"י 4-fold CV והאלמנט האחרון מחושב ע"י מודל בודד שמתאמן על כל מדגם האימון.

שימו לב כי בדומה לתרגיל בית 3, במימוש של הפונקציה `cross_validation_error` אסור לכם להשתמש בפונקציות עזר מהספרייה sklearn. בפרט אסור לכם להשתמש בפונקציה `cross_val_score` מתוך sklearn. עם זאת, אתם יכולים להשוות את הפלטים שלכם לפלטים של הפונקציות הרלוונטיות מהספריה. כמו כן, בפונקציה `SVM_results` מותר (ובדאי) להשתמש בפונקציות ומחלקות מ sklearn.

ד. חלקו את סט הנתונים לסט אימון וסט מבחן באמצעות הפקודה הבאה:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

הריצו את הפונקציה מסעיף ב' על הנתונים שטענתם בסעיף א'. ייתכן והריצה תיקח הרבה זמן (~שעה).

ציירו גרף עמודות (bar plot) המציג את התוצאות של כל ניסוי. כלומר, ציר ה-x יתאר את מודלי ה-SVM השונים שאימנם וציר ה-y יתאר את שגיאת האימון הממוצעת, שגיאת הולידציה הממוצעת ושגיאת המבחן (סה"כ 10 שלשות של עמודות). יש להקפיד על צבע שונה לכל סוג של עמודה (אימון / ולידציה / מבחן).

מיהו המודל הטוב ביותר לפי שיטת CV? מיהו המודל הטוב ביותר על מדגם המבחן? האם מדובר באותו המודל?