

התחיל ב:	09:00, 20/07/2022
מצב הסתיים	
הושלם ב-	11:00, 20/07/2022
הזמן שלקח	2 שעות
ציון	96.00 מתוך ציון מירבי של 100.00

שאלה 1

תקין

4.00 נקודות מתוך 4.00

השאלה מתייחסת ל-learning rate בשיטת linear regression ולא לגורמים הלמידה שלה, gradient descent

ה-learning rate האופטימלי הוא 0.5

אם ה-learning rate גבוה מידי, אזי gradient descent עלול להתכנס לאט מידי, ואילו אם ה-learning rate נמוך מידי, אזי gradient descent עלול לא להתכנס

אם ה-learning rate נמוך מידי או גבוה מידי, אזי gradient descent עלול להתכנס לאט מידי

אם ה-learning rate נמוך מידי, אזי gradient descent עלול להתכנס לאט מידי, ואילו אם ה-learning rate גבוה מידי, אזי gradient descent עלול לא להתכנס

אם ה-learning rate נמוך מידי או גבוה מידי, אזי gradient descent עלול לא להתכנס

אם ה-learning rate נמוך מידי או גבוה מידי, אזי gradient descent עלול להתכנס מהר מידי

שאלה 2

תקין

4.00 נקודות מתוך 4.00

להלן אמירות לגבי תהליך selection באלגוריתם אבולוציוני. אחת מהן אינה נכונה. סמנו אותה.

selection מייצר פרטים חדשים בעלי פיטנס טוב יותר

selection עובד אם ערכי הפיטנס שליליים

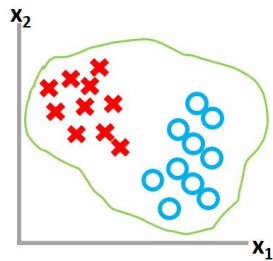
לצורך ביצוע selection ניתן למיין את האוכלוסייה ע"פ ערכי הפיטנס, אך אין חובה בכך

selection הינו הסתברותי

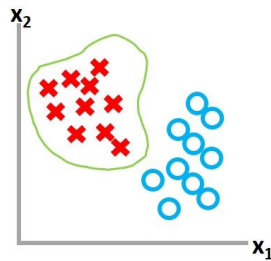
selection עובד אם ערכי הפיטנס מנורמלים או לא

selection אינו תלוי בייצוג פרט באוכלוסייה

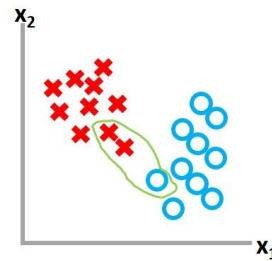
נתון dataset עם 2 classes, המסומנים ב-X אדום ו-O כחול, עבורו יש למצוא מודל המבצע קלסיפיקציה בינארית. 4 אנשים הציעו חלוקות שונות ל- training set (פריטים המוקפים בקו ירוק) ו- test set (היתר, שאינם מוקפים בקו ירוק). מי הציע את החלוקה הטובה ביותר?



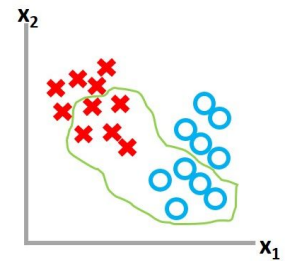
ג'יימי



סרסי



דאינריז



טיריון

ג'יימי

טיריון וגם ג'יימי

טיריון ☒

דאינריז וגם סרסי

דאינריז

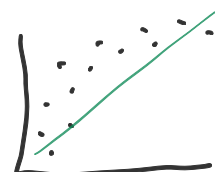
סרסי

שאלה 4

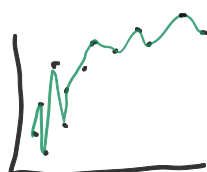
תקין

4.00 נקודות מתוך 4.00

אתם מריצים אלגוריתם למידת מכונה ורואים שה-error ב-training set נמוך, ואילו ה-error ב-test set גבוה.  
אילו מהמשפטים הבאים נכון:



High Bias  
Low Var



High Var  
Low Bias

אין bias ואין variance

יש לכם bias נמוך ו-variance גבוה ☒

האלגוריתם עובד מצוין

יש לכם bias נמוך ו-variance נמוך

יש לכם bias גבוה ו-variance גבוה

יש לכם bias גבוה ו-variance נמוך

שאלה 5

תקין

4.00 נקודות מתוך 4.00

מהי פונקציית אקטיבציה (activation function)?

פונקציה לא לינארית אשר משתמשים בה ב- Linear Regression

פונקציה לינארית אשר משתמשים בה ב- naive bayes

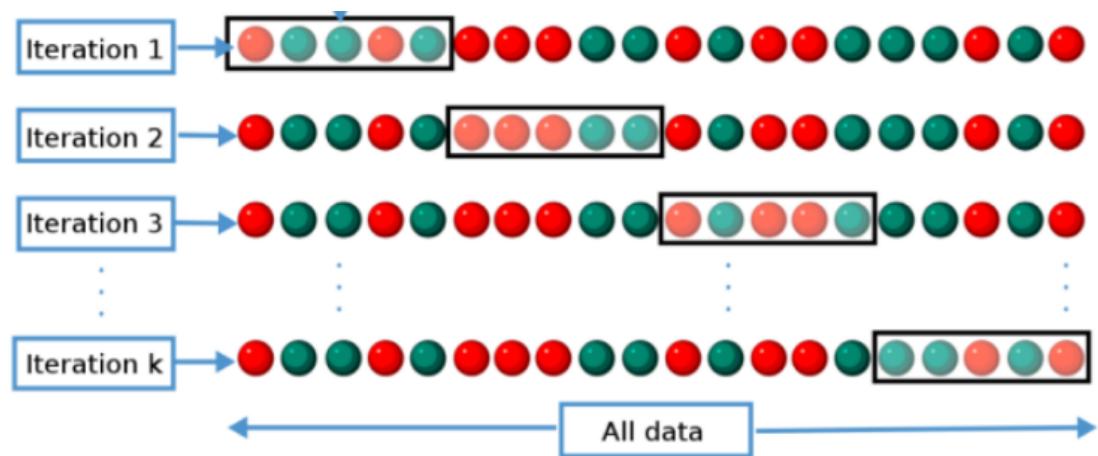
פונקציה לא לינארית אשר משתמשים בה ב- XGBoost

פונקציה לינארית אשר משתמשים בה ב- deep learning

כל 5 התשובות האחרות שגויות ☒

פונקציית הפעלה אשר משתמשים בה ב- Linear Regression

התמונה להלן ממחישה תופעה, תהליך, או אלגוריתם שראינו בכיתה. מהו?



data leakage (

data scaling (

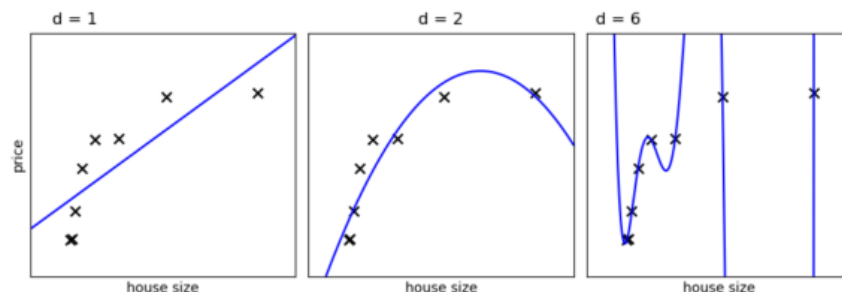
cross-validation 0 ✓

bias-variance tradeoff (

data normalization (

training vs. validation error (

עבור בעיית ניבוי מחיר בית כפונקציה של גודל הבית, אנו מבצעים 3 ריצות של רגרסיה פולינומילית. אחת עם מעלת פולינום  $d=1$ , שנייה עם מעלה  $d=2$ , שלישית עם מעלה  $d=6$ . מהו התיאור המדויק ביותר של התוצאות, המוצגות בגרף שלהלן?



בכל שלושת המקרים יש לנו under-fitting

עבור  $d=1$  יש לנו מקרה של over-fitting ועבור  $d=6$  יש לנו מקרה של under-fitting

בכל שלושת המקרים יש לנו over-fitting

עבור  $d=1$  יש לנו מקרה של under-fitting ועבור  $d=6$  יש לנו מקרה של over-fitting

עבור  $d=1$  ו- $d=2$  יש לנו מקרה של over-fitting ועבור  $d=6$  יש לנו מקרה של under-fitting

עבור  $d=1$  ו- $d=2$  יש לנו מקרה של under-fitting ועבור  $d=6$  יש לנו מקרה של over-fitting

ברוב ה-classifiers וה-regressors ב-scikit-learn ישנה פונקציית fit. מה תפקידה?

להפריד דאטא ל-train ו-test, לאמן את המודל על train, ולהחזיר מודל מאומן

לחשב את פונקציית הפיטנס של המודל

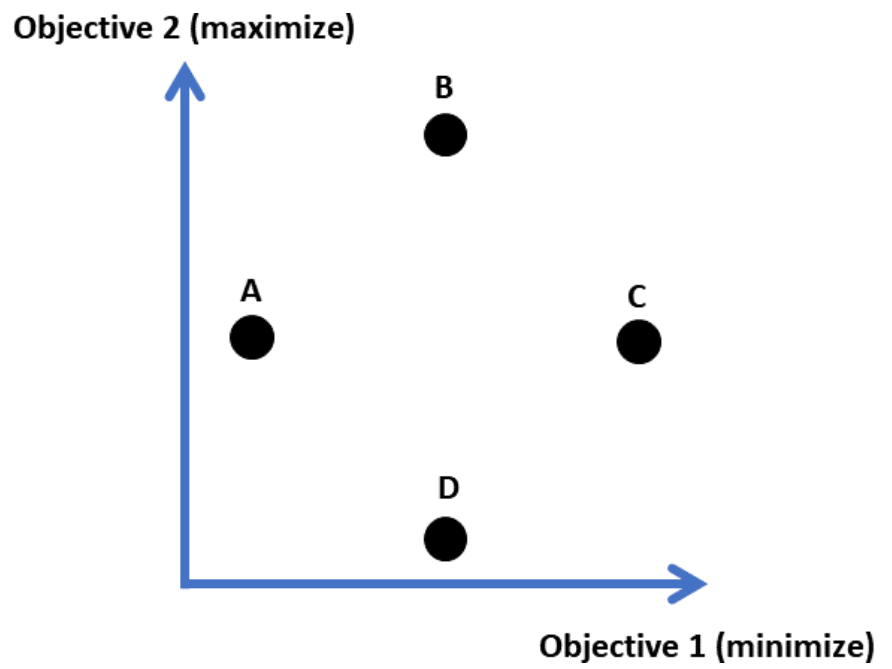
להפריד דאטא ל-train ו-test, לאמן את המודל על train, ולבדוק על test

לאמן את המודל על סמך נתוני אימון

לבדוק האם פונקציית הפיטנס סובלת מבעיה של התכנסות מוקדמת

לבדוק את ההתאמה של המודל לנתוני הבדיקה (test)

אילו נקודות שייכות לחזית ה-Pareto, כפי שהגדרנו בכיתה?



A,D

A

A,C

A,B

A,B,C

B,C



שאלה 10

תקין

4.00 נקודות מתוך 4.00

להלן רשימת קביעות לגבי fitness-proportionate selection. אחת מהן אינה נכונה. איזו?

0

יכול לשנות פתרון טוב לרעה

פתרון בעל פיטנס גבוה עלול להשתלט על האוכלוסיה

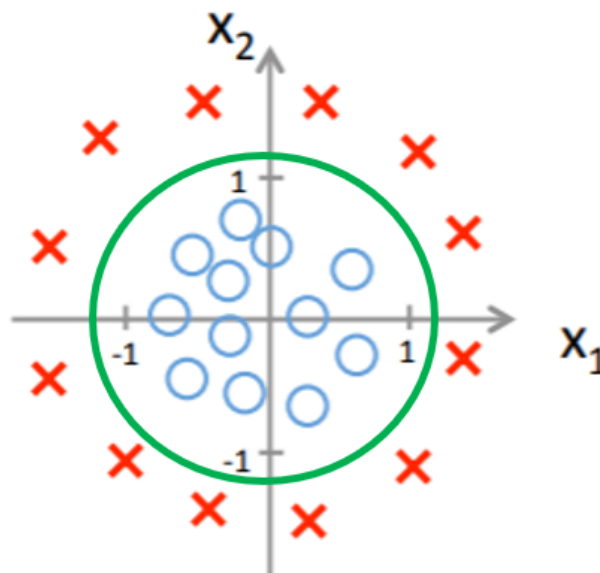
דורש חישוב הסתברויות יחסיות של כל פרט

פתרון בעל פיטנס גבוה עלול לא להבחר

עלול לא לעבוד טוב כאשר יש פתרונות בעלי פיטנס דומה לאחר דורות רבים

חשוף ל-function transposition

האם logistic regression יכול למצוא פתרון עם decision boundary לאחד מ-2 classes (עיגול כחול או איקס אדום)?



- כן, אם נשתמש ב-polynomial features עם פולינום ממעלה 1
- כן, אם נגדיל את המספר המינימאלי של samples הנחוצים ל-split בכל צמת במודל
- לא, מאחר ו-logistic regression אינו מתכנס כאשר ה-decision boundary הינו עיגול
- לא, מאחר ו-logistic regression מייצר פתרונות עבורם ה-decision boundary הינו קו ישר ולא עיגול
- כן, אם נשתמש ב-polynomial features עם פולינום ממעלה אי זוגית
- כן, אם נשתמש ב-polynomial features עם פולינום ממעלה 2

שאלה 12

תקין

4.00 נקודות מתוך 4.00

מנה 3 הבדלים בין random forest לבין AdaBoost

- a. עץ מלא מול עץ לא מלא, אקראיות שמקורה יוניפורמי מול גאוסיאני, חשיבות לסדר העצים מול אי חשיבות לסדר
- b. עץ AVL מול עץ בינארי, משקל שווה מול משקל לא שווה של העצים, חשיבות לסדר העצים מול אי חשיבות לסדר
- c. עץ מלא מול גדם, משקל שווה מול משקל לא שווה של העצים, חשיבות לסדר העצים מול אי חשיבות לסדר
- d. עץ מלא מול גדם, משקל שווה מול משקל לא שווה של העצים, בעיות קלספיקציה בלבד מול בעיות קלספיקציה+רגרסיה
- e. עץ מלא מול גרף, משקל שווה מול משקל לא שווה של הגרפים, מיון מול אי מיון העצים
- f. גרף חסר מעגלים מול עץ מלא, משקל שווה מול משקל לא שווה של העצים, חשיבות לסדר העצים מול אי חשיבות לסדר

שאלה 13

תקין

4.00 נקודות מתוך 4.00

מה ההבדל בין gradient descent (GD ל-stochastic gradient descent (SGD?


- a. ל-GD התכנסות מובטחת ואילו ל-SGD לא
- b. SGD משתמש בחלק מהדאטא ו-GD משתמש בכל הדאטא
- c. GD משתמש בחלק מהדאטא ו-SGD משתמש בכל הדאטא
- d. SGD הינו הסתברותי ואילו GD הינו דטרמיניסטי
- e. GD הינו הסתברותי ואילו SGD הינו דטרמיניסטי
- f. ל-SGD התכנסות מובטחת ואילו ל-GD לא

שאלה 14

תקין

4.00 נקודות מתוך 4.00

אנו נרצה להשתמש ב ridge regression או lasso regression לעומת הוורסיה הבסיסית, קרי linear regression, כי:


- a. ל ridge ו-lasso יש אקטיבציות לינאריות שפועלות כמנגנוני תיקון שגיאות
- b. ל ridge ו-lasso יש מנגנוני רגולריזציה להתמודדות עם משקלים נמוכים
- c. ל ridge ו-lasso יש מנגנוני רגולריזציה להתמודדות עם משקלים גבוהים 
- d. ל ridge ו-lasso יש יתרונות מהירות משמעותיים
- e. ל ridge ו-lasso יש מנגנוני רגולריזציה להתמודדות עם משקלים קרובים לאפס
- f. ל ridge ו-lasso אין דרישה ל-scaling

שאלה 15

תקין

4.00 נקודות מתוך 4.00

ברגרסיה לינארית מתי ה sum of squared residuals שווה ל-0?

- a. כאשר כל נקודות ה-training מנורמלות לערכים בתחום:  $[-1, 1]$
- b. כאשר אין outliers ב-דאטא
- c. כאשר כל נקודות ה-training מנורמלות לערכים בתחום:  $[0, 1]$
- d. כאשר כל נקודות ה-training על קו ישר אחד 
- e. כאשר כל הדאטא מתפלג יוניפורמית
- f. כאשר כל הדאטא מתפלג נורמלית

שאלה 16

תקין

4.00 נקודות מתוך 4.00

מהם 2 מקורות לאקראיות ב-random forest?

- a. random graphs, random features
- b. random builds, random trees
- c. random trees, random feature subset
- d. random classes, random feature subset
- e. bootstrapped trees, bootstrapped features
- f. dataset bootstrapping, random subset of features

שאלה 17

תקין

4.00 נקודות מתוך 4.00

לאחר הרצה של k-fold cross-validation באיזה מודל נשתמש לניבוי?

- a. כל התשובות האחרות שגויות
- b. k המודלים הכי טובים
- c. ניתן להשתמש בכל אחד מ-k המודלים
- d. ממוצע של כל k המודלים
- e. k-1 המודלים הכי טובים
- f. המודל עבורו קבלנו את ה-score הכי טוב

שאלה 18

שגוי

0.00 נקודות מתוך 4.00

מהי הנחה בסיסית של אלגוריתם Naive Bayes?

- a. אי תלות של features ✓
- b. אי תלות של classes
- c. אי תלות של פרמטרי איתחול
- d. אי תלות של data
- e. אי תלות של weights
- f. אי תלות של התפלגויות

מה עושה הקוד הבא:

```
import torch
import optuna

def objective(trial):

    n_layers = trial.suggest_int('n_layers', 1, 3)
    layers = []

    in_features = 28 * 28
    for i in range(n_layers):
        out_features = trial.suggest_int(f'n_units_l{i}', 4, 128)
        layers.append(torch.nn.Linear(in_features, out_features))
        layers.append(torch.nn.ReLU())
        in_features = out_features
    layers.append(torch.nn.Linear(in_features, 10))
    layers.append(torch.nn.LogSoftmax(dim=1))
    model = torch.nn.Sequential(*layers).to(torch.device('cpu'))
    ... # stuff we don't need for this question
    return accuracy

study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=100)
```

- a. קונבולוציה + max pooling
- b. חישוב מספר ה out\_features בהתאם למשתנה in\_features
- c. ☒ מציאת ארכיטקטורת deep network
- d. סיווג למספר classes השווה ל- n\_units
- e. מימוש יעיל של אלגוריתם backpropagation ב-deep learning כאשר מספר האיטרציות שווה ל n\_trials
- f. קונבולוציה

למה נועד הקוד הבא?

```
from sklearn import datasets
iris = datasets.load_iris()
X = iris.data
y = iris.target

from sklearn.cluster import KMeans
clust = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
    kmeans.fit(X)
    clust.append(kmeans.inertia_)

from matplotlib import pyplot as plt
plt.plot(range(1, 11), clust)
plt.show()
```

בחירת ה-k הכי טוב

a

b. הרצת k-means ולאחריו חישוב השכנים כחלק מ-k nearest neighbors

c. מציאת שיטת האתחול הנכונה עבור k-means

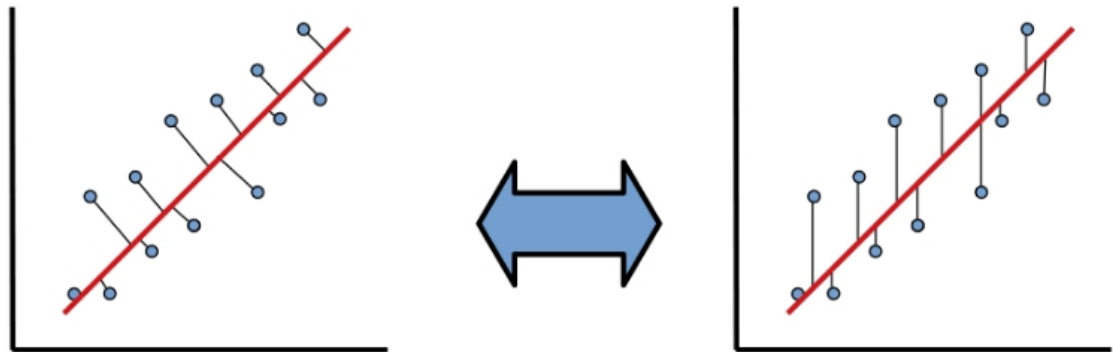
d. חישוב ה-clusters והצגתם

e. שילוב אופטימלי של k ו-knn

f. זהו קוד שגוי מאחר ויש קריאות ל-fit אבל אין קריאות ל-predict



התמונה הבאה, שראינו באחד השיעורים, ממחישה מה מול מה?



- a. Logistic Regression vs. Deep Network
- b. PCA vs Deep Network
- c. PCA vs Linear Regression
- d. Linear Regression vs. Logistic Regression
- e. Random Forest vs. Decision Tree
- f. XGBoost vs. AdaBoost

שאלה 22

תקין

4.00 נקודות מתוך 4.00

סמן את האימרה השגויה

- a. ה-permutation test מעבד תתי קבוצות בדאטא
- b. ה-permutation test "מערבב" את שני ה-samples
- c. ה-t-test מניח התפלגות נורמלית של שני ה-samples
- d. ה-t-test מניח random sampling של ה-samples
- e. ה-permutation test מניח התפלגות נורמלית של שני ה-samples
- f. ה-t-test מניח אי תלות בין שני ה-samples

שאלה 23

תקין

4.00 נקודות מתוך 4.00

השלם:

פונקציית ה-ArgMax בעייתית בהפעלת כלל השרשרת ב deep learning כי

- a. הנגזרת שלילית
- b. לא ניתן לעבוד עם ArgMax בשיתוף עם SoftMax
- c. הפונקצייה יקרה מאד לחישוב כאשר עובדים על GPU
- d. ה-SoftMax שבא לאחר ה-ArgMax "הורס" את ערכה
- e. ה-SoftMax שבא לפני ה-ArgMax "הורס" את ערכה
- f. הנגזרת היא 0

זהו קוד ממאמר שלי שראינו בכיתה.

על איזה מודל הוא מבוסס?

```
def fit(self, X, y):  
    y_res = deepcopy(y)  
    for model in self.models:  
        model.fit(X, y_res)  
        p = model.predict(X)  
        y_res -= np.nan_to_num(p)  
    return self  
  
def predict(self, X):  
    pred = np.zeros(X.shape[0])  
    for model in self.models:  
        pred += np.nan_to_num(model.predict(X))  
    return np.nan_to_num(pred)
```

gradient boosting ☒ a

ridge regression .b

adaptive boosting .c

stochastic gradient descent .d

gradient descent .e


random forest .f

שאלה 25

תקין

4.00 נקודות מתוך 4.00

לאחר שהרצנו אלגוריתם למידה כלשהו על ה- training set וקבלנו מודל שהוא fitted בצענו בדיקה על ה- test set וחישבנו מדד ROC-AUC של 1.02. המשמעות היא:

- a. לא ניתן להשיב מאחר ולא יודע אחוז ה-split בין ה-train ל-test (
- b. זהו מודל מצוין לאור המדד הגבוה על ה-test set (
- c. לא ניתן לדעת מאחר ולא נאמר אם הערך הוא scaled או לא (
- d. לא ניתן לדעת מאחר ולא נאמר אם הערך מנורמל או לא (
- e. זוהי שגיאה 
- f. זהו מודל חלש מאחר והוא רחוק מהשגיאה המינימאלית שהיא 0 (