בית ספר "אפי ארזי" למדעי המחשב המרכז הבינתחומי
The Efi Arazi school of computer science
The Interdisciplinary Center

סמסטר ב' תש"פ
Spring 2020

מבחן מועד א בלמידה ממוכנת
Machine Learning Exam A

Lecturer: Prof Zohar Yakhini
Time limit: 2 hours

מרצה: פרופ זהר יכיני
משך המבחן: 2 שעות

You should answer Qn 1 (mandatory), for 40 pts and two out of the other three questions (30pts each, 60pts together).

In the first page indicate the numbers of the two questions you answered. If there is no indication then the first two solutions will be graded.

Good Luck!

יש לענות על שאלה מספר 1 (חובה), 40 נקודות, ולבחור שתים מתוך שלוש השאלות הנוספות (30 נקודות לכל אחת, סה"כ 60 נקודות).

בעמוד הראשון יש לרשום את מספרי שתי שאלות הבחירה שעליהן בחרת לענות. אם לא יהיה רשום תיבדקנה שתי התשובות הראשונות.

בהצלחה!

You can use all Moodle course materials as well as student personal notes prepared before the exam. There is no need to print the material. You can use the appropriate files on your PC.
You can use calculators.
Justify all your answers and show your calculations.
All answers should be legibly written and scanned for submission as per IDC's instructions.

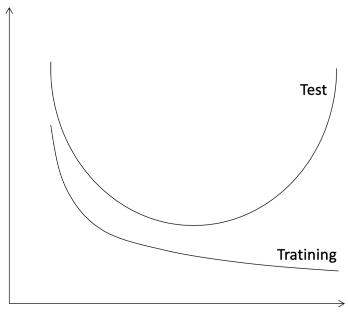
ניתן להשתמש בכל החומר הקיים ב Moodle ובנוסף סיכומי סטודנטים שנכתבו לפני תחילת המבחן. אין צורך להדפיס את החומר. אפשר להשתמש במחשב כדי לגשת לקבצים הרלוונטיים.
ניתן להשתמש במחשבון.
יש להצדיק את כל תשובותיך ולהראות את צעדי החישוב.
כל התשובות צריכות להיות כתובות בכתב ברור וקריא ולהיסרק להגשה ע"פ הנחיות IDC.

Question 1 (3 parts, 40 points) – MANDATORY QUESTION

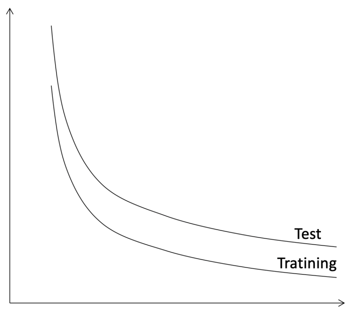
A. You receive 10,000 samples of labeled data, split into training and test, to perform regression to predict house prices from 4 real valued features (x_1, x_2, x_3, x_4).
   1. (4 points) If you preform Linear Regression with all 4 features, what is the dimension of the inferred vector of coefficients $\vec{\theta}$?
   2. (5 points) If you preform Polynomial Regression using <u>all</u> monomials of degree less or equal 5
      (for example: $x_1, x_1^2, x_1^3, x_1^4, x_1^5, x_1^2x_2^2, x_2x_3, x_1^2x_2^2x_3, x_4^3x_3^2, x_3^3$).
      What is the dimension of the inferred vector of coefficients $\vec{\theta}$?
   3. (5 points) You performed Polynomial Regression using all monomials of degree less or equal 2. You obtained an MSE on both training and test with which you are not happy. Your colleague proposes that you also add $2x_1x_2$ and $2x_3x_4$ to your vector of features since these are important quantities in the context of pricing houses.
      Will this change help the MSE performance? Explain.
   4. (6 points) Assume that your training / test split was 8,000 / 2,000. You perform Polynomial Regression of increasing degrees.
      For each of the following graphs state whether it can conceivably describe the observed performance.
      In all graphs below the y-axis represents the MSE and the x-axis represents the degree of the Polynomial Regression.
      Explain.

A

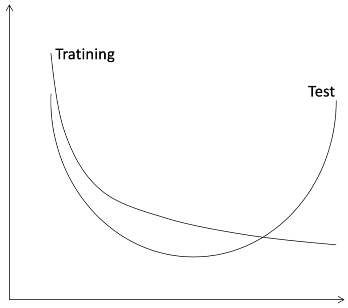
B



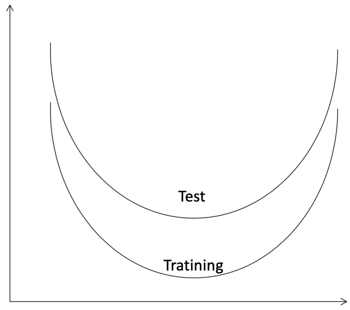
C



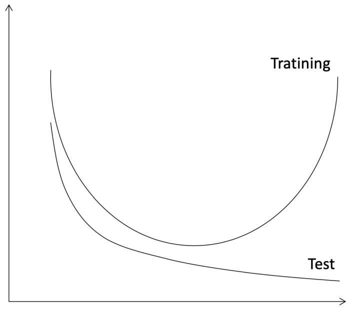
D



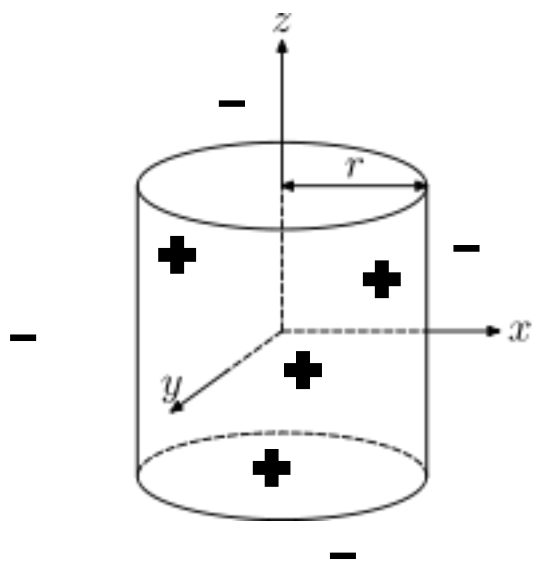
E



F

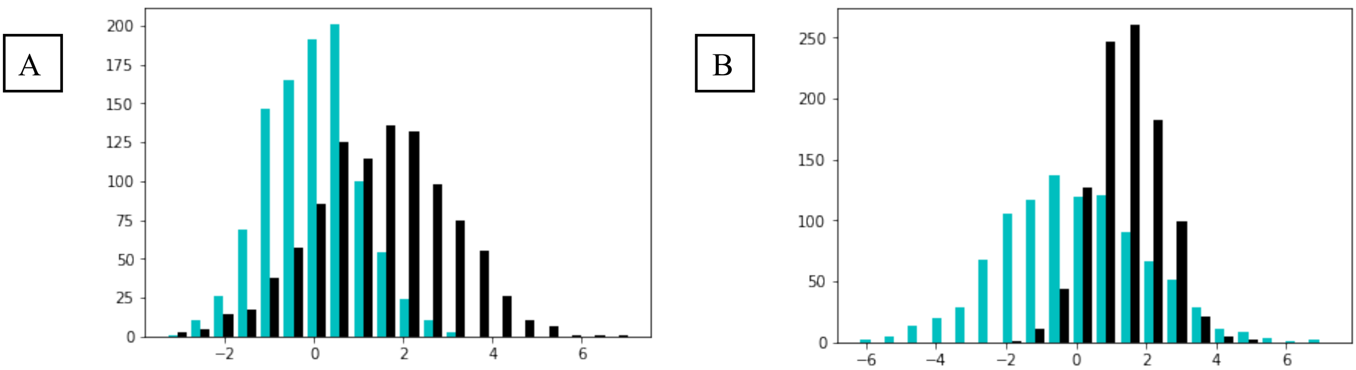


B. (10 points) Let $X = \mathbb{R}^3$. Calculate the VC dimension of the set of all cylinders threaded on the z axis. The cylinder can be both on the positive and negative part of z. Consider data points inside the cylinder as Positives and data points outside of the cylinder classify as Negatives.
Formally:
For any three numbers $l, u \in \mathbb{R}, r \in \mathbb{R}_+$
we define $h(l,u,r) = \{(x,y,z) \mid \ x^2 + y^2 \leq r \ \wedge \ \ l \leq z \leq u\}$.
And now we define the hypotheses space as:
$H = \{h(l,u,r) \mid \ l,u \in \mathbb{R}, r \in \mathbb{R}_+\}$

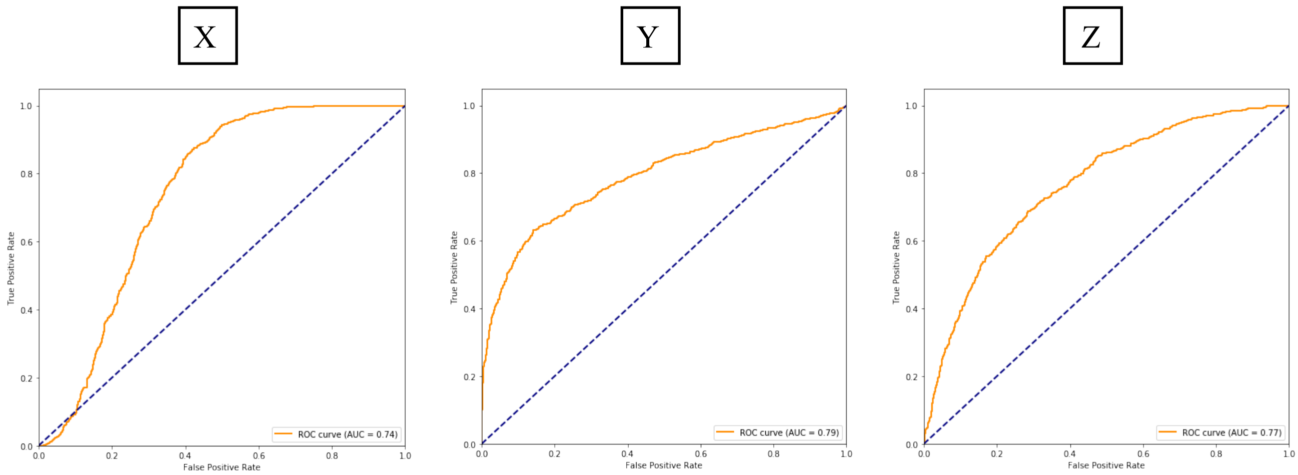


C. Two companies are proposing a Corona test to a hospital. The hospital asks for your advice as to which test is better. The plots below represent the distributions of the tested quantity for the two companies. The black histograms represent the positive cases. We further depict the ROC curves computed to evaluate their performance.

Distributions



ROC curves



1. (4 points) Match two of the three ROC curves above, X, Y and Z, with the two distributions, A and B. Explain.

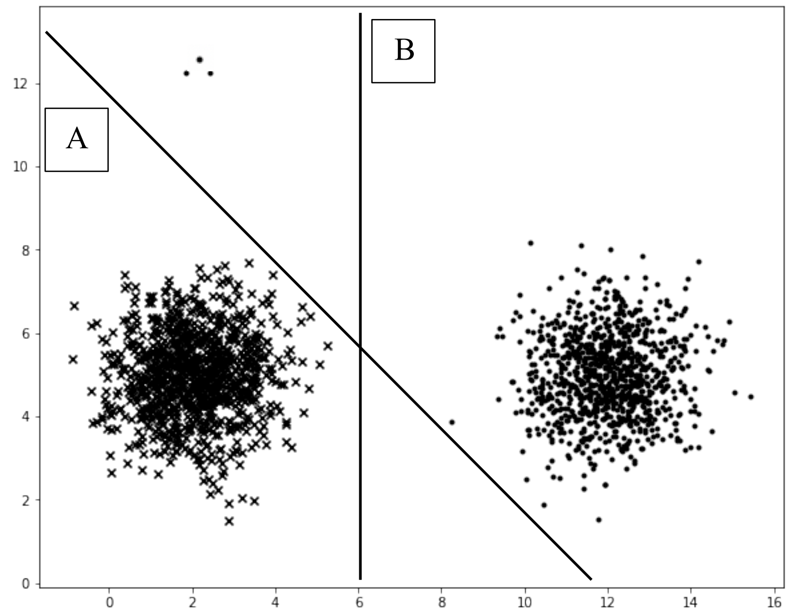
Recall that the expected benefit of a test, for the hospital, is measured by

$$\pi = \alpha * TPR - FPR$$

2. (3 points) Assuming $\alpha = 23$ which company will you select?
3. (3 points) Assuming $\alpha = 0.5$ which company will you select?

Question 2 (6 parts 30 points)

A. (5 points) Consider the data in the figure below and the two different linear decision boundaries (A & B) as depicted.
One of them was obtained by running Logistic Regression on the data and the other one was obtained by running Naïve Bayes with normal distributions.
Which is which? Explain.



The following pertains to parts B through F.
In rolling two dice, the first with 6 sides and the second with 4 sides, we have the following distributions of results for two casino houses – Casino A and Casino B:

Casino A

Die1 \ Die2	1	2	3	4
1	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
2	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
3	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
4	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
5	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$
6	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$

Casino B

Die1 \ Die2	1	2	3	4
1	$\frac{1}{12}$	$\frac{1}{24}$	0	$\frac{1}{24}$
2	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{24}$	0
3	0	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{24}$
4	$\frac{1}{24}$	0	$\frac{1}{24}$	$\frac{1}{12}$
5	$\frac{1}{12}$	0	$\frac{1}{12}$	0
6	0	$\frac{1}{12}$	0	$\frac{1}{12}$

The prior probability for playing in Casino A is $\frac{2}{5}$.

Given a game outcome (2 numbers), we want to classify whether the game was played in Casino A or B.

B. (5 points) We observe the following game outcome: 1^{st} die is 6, 2^{nd} die is 2. Which casino will a Naïve Bayes classifier predict? Show your calculations.

C. (5 points) Given the same game result as in Part B, which casino will a Full Bayes classifier predict? Show your calculations.

D. (5 points) What is the minimal prior we need to assign to Casino A in order for the Full Bayes classifier to predict A regardless of the game outcome?
That is, find the minimal number π that satisfies:
(Prior of A > π) $\Longrightarrow$ (Full Bayes always decides A).
Show/explain your calculations.

E. (5 points) Assume that the prior of A is the number π that you found in Part D. You can now change two entries in the joint distribution matrix of Casino B. Perform a change that produces a joint distribution that will lead the Full Bayes classifier to select Casino B, when observing the results (6,2), like in Part B.

F. (5 points) What is the minimal prior we need to assign to Casino A in order for the task of part E to be impossible? Explain

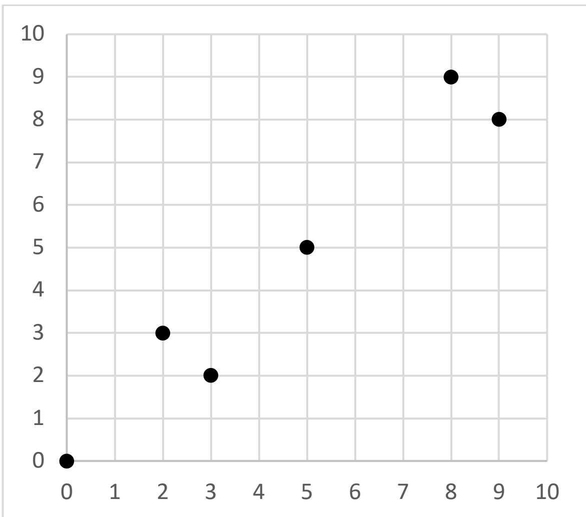
Question 3 (5 parts 30 points)

A. (10 points) We want to cluster a set of S instances into k groups. Below is a pseudo code of an algorithm called *k-medians* with L_1 as the distance measure:

Initialize $c_1, \dots, c_k$ by randomly.
Loop:
  Assign all n instances to their closest c_i, with L_1 as the distance metric, and create k clusters $S_1, \dots, S_k$
  For each cluster S_i $(1 \le i \le k)$ define a new c_i:
    $c_i = \text{median}(S_i)$
Until no change in $c_1, \dots, c_k$
Return $c_1, \dots, c_k$

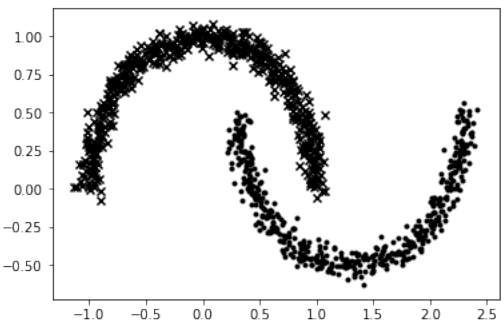
* Note: the median of a set of vectors $S = v_1, \dots, v_n$ is a vector of the same dimension. Each entry of this vector is obtained by computing the median of the corresponding entries in $v_1, \dots, v_n$.

Consider a set of instances, S, of size 6 with 2 features as shown in the figure and the table. Run the k-medians algorithm with $k = 2$ and with the starting centers at $c_1 = (7, 9)$ and $c_2 = (7,8)$.
At each iteration write down the new centers and which center each point is assigned to. No need to show all intermediate calculations.



x	y
0	0
2	3
3	2
5	5
8	9
9	8

B. (5 points) Consider the following cluster structure. The 'xs' (the upper half circle) belong to Cluster 1 and the 'dots' (the lower half circle) belong to Cluster 2 (see picture).
Can it be the final assignment of the execution of k-medians? Explain.



C. (5 points) Give data for which k-medians, as described above, converges to a different clustering assignment solution than the one resulting from running k-means, starting at the same initial points, with k=2.
Justify your answer.

D. (5 points) We ran k-means with k=8 on the following image for color quantization:

We ran the algorithm twice.
We got the following two images:

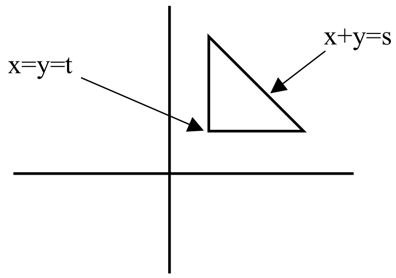
Explain how it is possible for us to get two different images.

E. (5 points) How many parameters do you need to learn in order to infer a multivariate Gaussian mixture model in $\mathbb{R}^2$ with $k = 3$?

Question 4 (3 parts 30 points)

A. (10 points) Prove:
   If $\alpha, \beta > 0$ and K, L are kernels then $\alpha K + \beta L$ is also kernel.

B. Consider the instance space $X = \mathbb{R}^2$ with some probability distribution π.
   Consider the concept space C that consists of right-angle isosceles triangles whose head vertex is positioned on the $x = y$ line and whose base is to the right of its head vertex.
   Formally, for any $t \geq 0$ and $s > t$ define
   $c(s,t) = \{(x,y) |\, x \geq t,\ y \geq t\,,\ x + y\ \leq s\}$
   and then:
   $C\ = \{c(s,t) | s > t \geq 0\}$.
   Let $H = C$.



1. (6 pts) Propose a consistent learning algorithm L that takes as input labeled points $D^m = \{(x_1, y_1), (x_2, y_2) \dots, (x_m, y_m)\} \subseteq \mathbb{R}^2$ and returns $h \in H$.
2. (7 pts) Prove that C is PAC learnable from H by computing a sufficiency bound on the sample complexity.
3. (7 pts) For $\varepsilon = 0.05$ and $\delta = 0.01$ compute a sufficiency bound on the size, m, of a set D^m of independently drawn training samples, that guarantees that for any $c \in C$ we have:
   $Prob\big(Err\big(c, L(D^m)\big) > \ \varepsilon\big) < \delta$

GOOD LUCK!