מספר מחברת: ________

ת.ז: ______________

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

סמסטר ב' תשפ"א
Spring 2021

מבחן מועד א בלמידה ממוכנת
Machine Learning Exam A

Lecturer: Prof Zohar Yakhini
Time limit: 3 hours

מרצה: פרופ זהר יכיני
משך המבחן: 3 שעות

Answer 4 out of 5 from the following questions. Each question is 25 points.

Good Luck!

יש לענות על 4 מתוך 5 השאלות הבאות.
לכל השאלות משקל שווה (25 נקודות)

בהצלחה!

ניתן להשתמש בדפי העזר המצורפים, מחשבון ומילון בלבד. כל חומר עזר אחר אסור.

יש להסביר\להוכיח את כל התשובות.

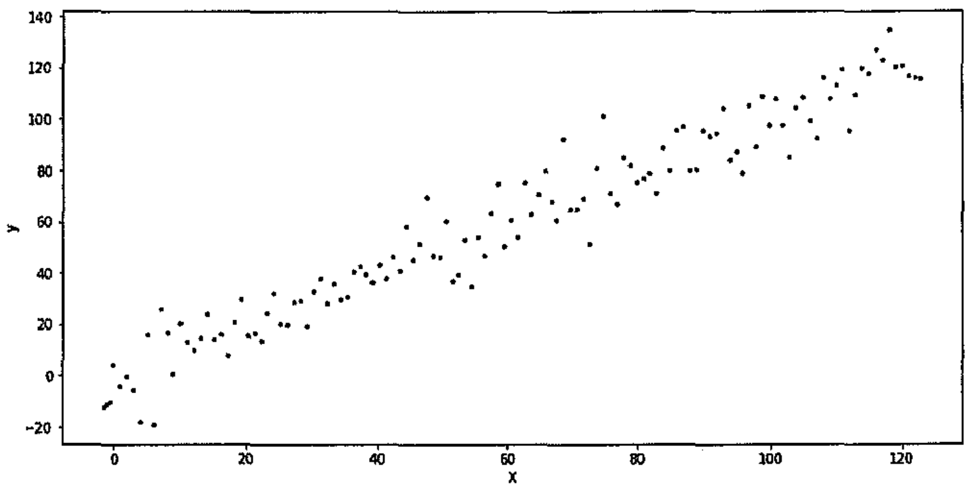
You can use the attached formula sheet, a calculator and a dictionary. All other material should not be used.

Prove/explain all your answers.

לתשומת לבך, אין לכתוב
בצידו השני של הדף!
אין להשתמש במרקר!

שאלה 1 – רגרסיה ליניארית (25 נקודות)

1. נתונה קבוצת הנקודות הבאה:



מטרתנו היא לבדוק איזה אלגוריתם יכלול בצורה טובה יותר, בניסיון לחזות את y באמצעות X:
רגרסיה ליניארית או 3-NN (אלגוריתם k-NN עם $k = 3$).

הצעת הקולגה שלך היא להשתמש ב-Cross Validation בדרך הבאה: לחלק את הנקודות ל-3 קבוצות: [-5,40), [40, 80), [80,125] ולהריץ *Cross Validation*, ראשית עם רגרסיה לינארית ושנית עם 3-NN, ולבסוף לבחור את האלגוריתם עם ה-MSE הממוצע הנמוך יותר.

a. (4 נק') איזה אלגוריתם יבחר? הסבר.
b. (4 נק') האם ישנה בעיה בגישה של הקולגה? הסבר.
c. (4 נק') האם ניתן למצוא 3 נקודות חדשות (נקודות test) שהתוצאה עליהן תהיה טובה יותר כאשר משתמשים ב-3-NN? ציין את ערכי הנקודות. הסבר.

2. נתונה מטריצת דוגמאות X בעלת מימד $m \times n$ (m דוגמאות ו-n פיצ'רים).
   a. (3 נק') ברגרסיה לינארית סטנדרטית, מכמה מקדמים בנוי המודל?
   b. (3 נק') ברגרסיה פולינומיאלית עם מעלה r, מכמה מקדמים בנוי המודל? (ברגרסיה לינארית $r = 1$).

3. רגרסיה לינארית, העושה שימוש בפונקציית הפסד MSE, מנסה למצוא:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \; ||X\theta - y||_2^2$$

   a. (2 נק') איך היית משנה את ההגדרה של פונקציית ההפסד MSE, כך שכל דוגמא יכולה לקבל משקל שונה, w_i, בחישוב ההפסד.
   b. (5 נק') מצא מטריצה W שבעזרתה תוכל לעדכן את משוואת הרגרסיה הלינארית ולהגדיר פתרון Pseudo Inverse עבור הפונקציה שהגדרת לעיל.
   הסבר את כל הצעדים. בתשובתך ציין מהי המטריצה ואת משוואת המינימיזציה המעודכנת.

שאלה 2 – עצי החלטה (25 נקודות)

1. (10 נק') אנו מנסים לבצע סיווג בין בני אדם (H) לבין אנשי מאדים (M) באמצעות הפיצ'רים הבאים: $Height \in \{S,T\}$, $Legs \in \{2,3\}$, $Green \in \{N,Y\}$ ו-$Smelly \in \{N,Y\}$.

באמצעות הדאטה בטבלה, כאשר הזן אותו אנו מנסים לסווג מופיע בעמודה Species, בנה את שתי הרמות הראשונות (פצלו את השורש ואת צאצאיו של השורש) של עץ החלטה באמצעות אנטרופיה והאלגוריתם החמדן שנלמד בכיתה.
ציירו את העץ שקיבלתם.
האם נדרש לבצע פיצול נוסף על מנת לקבל עץ החלטה עם שגיאת אימון אפס?

Green	Legs	Height	Smelly	Species
N	3	S	Y	M
Y	2	T	N	M
Y	3	T	N	M
N	2	S	Y	M
Y	3	T	N	M
N	2	T	Y	H
N	2	S	N	H
N	2	T	N	H
Y	2	S	N	H
N	2	T	Y	H

2. (5 נק') באמצעות אותו דאטה המופיע בסעיף (1), רשמו עץ החלטה בגובה 2 עם שגיאת אימון אפס (עלים מונוכרומטיים). רמז: ניתן להתעלם מחלק מהפיצ'רים.
3. (3 נק') האם ניתן להחליף את התפקיד של Chi-Sqaure ואנטרופיה בעת בניית וגיזום עץ החלטה? למשל, האם ניתן להשתמש ב-Chi-Sqaure על מנת לבחור את הפיצ'ר הטוב ביותר לפיו נבצע פיצול בעץ ולהשתמש באנטרופיה על מנת לבצע גיזום. הסבירו.
4. (7 נק') לפניכם שני עצי החלטה:
   a. העץ הראשון אומן על דאטה כלשהו, $X_{m\times n}$, באמצעות אנטרופיה.
   b. אינדקסים נדגמו בצורה רנדומלית מהתפלגות בינומית $Bin(0.5, n)$. תהי S קבוצת האינדקסים שנדגמו. נגדיר מטריצה אלכסונית A בגודל $n \times n$ כך ש-$A_{i,i} = 1$ אם ורק אם האינדקס $i \in S$ ו-$A_{i,i} = -1$ אחרת. תהי $Y = X \cdot A$. העץ השני אומן על דאטה Y באמצעות אנטרופיה.

   האם לשני העצים יש את אותן הפרדיקציות על כל הדוגמאות בסט האימון? הסבירו\הוכיחו את תשובתכם.

שאלה 3 – תיאוריה (25 נקודות)

נתון $X = \mathbb{R}^3$. תהי $C = H$ קבוצת כל הצילינדרים שהציר המרכזי שלהם נמצא על ציר ה-z והמרכז שלהם בראשית הצירים. נקודות שנמצאות בתוך הגליל מסווגות כחיוביות ונקודות מחוץ לגליל כשליליות.

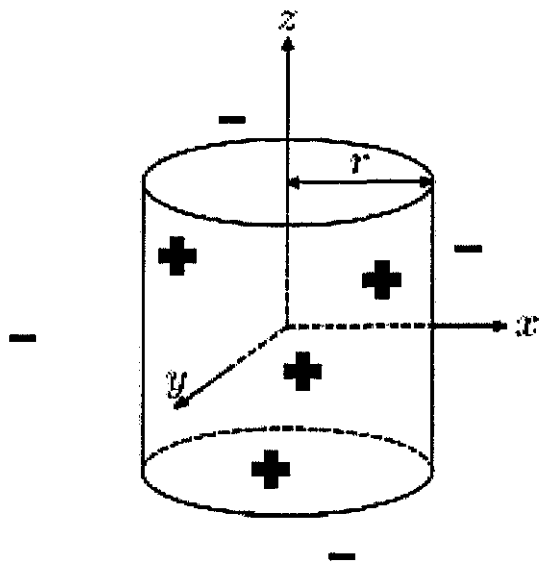
פורמלית:

לכל שני מספרים $l, r \in \mathbb{R}_+$

נגדיר $h(l,r) = \{(x,y,z) \mid \; x^2 + y^2 \le r \wedge -l \le z \le l\}$.

וכעת נגדיר את מרחב ההיפותזות: $H = \{h(l,r) \mid l, r \in \mathbb{R}_+\}$.

תהי π התפלגות כלשהי על X.



1. (5 נק') חשב את מימד ה-VC של H. הוכיחו את תשובתכם.
2. (6 נק') הצע אלגוריתם עקבי (consistent) L שמקבל כקלט נקודות מתויגות $D^m = \{(x_1,y_1,z_1),(x_2,y_2,z_2),\dots,(x_m,y_m,z_m)\} \subseteq \mathbb{R}^3$ ומחזיר $h \in H$.
3. (7 נק') חשב חסם מספק על מורכבות המדגם (sample complexity) בלמידה של C ע"י H בשימוש באלגוריתם L שהצעת. כלומר, בהינתן (ε, δ), חשב חסם על גודל המדגם m, של קבוצה D^m של דוגמאות אימון, שנדגמו בצורה בלתי תלויה, המבטיח שעבור כל $c \in C$ יתקיים:

$$Prob\big(Err(c, L(D^m)) > \varepsilon\big) < \delta$$

4. (7 נק') עבור $\varepsilon = 0.05$ ו-$\delta = 0.01$ חשבו שני חסמים מספקים: אחד באמצעות החישוב שביצעתם בסעיף הקודם והשני באמצעות מימד ה-VC. העזרו בנוסחא הבאה לחישוב ה-Sample Complexity באמצעות מימד ה-VC כאשר $C = H$:

$$m \ge \frac{1}{\varepsilon}\left(4\log_2\left(\frac{2}{\delta}\right) + 8VC(H)\log_2\left(\frac{13}{\varepsilon}\right)\right)$$

איזה חסם הדוק יותר?

שאלה 4 – Bayes ו-MLE (25 נקודות)

1. (5 נק') בהינתן דוגמאות מעל הקלאסים $(A_1, A_2, \dots, A_r)$ והפיצ'רים $(x_1, x_2, \dots, x_n)$, כתבו את נוסחאות הסיווג עבור:
   a. Naïve Bayes
   b. Full Bayes

בהטלת שתי קוביות בעלות 6 פאות כל אחת, התקבלו ההתפלגויות הבאות עבור שני בתי קזינו: Casino Golden Peacock (CGP) ו-Casino Silver Turkey (CST).

Casino Golden Peacock

קוביה1 \ קוביה 2	1	2	3	4	5	6
1	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0
2	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0
3	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0	0
4	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$	0
5	0	0	0	0	$\frac{1}{12}$	$\frac{1}{12}$
6	$\frac{1}{12}$	0	0	0	0	$\frac{1}{12}$

Casino Silver Turkey

קוביה1 \ קוביה 2	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

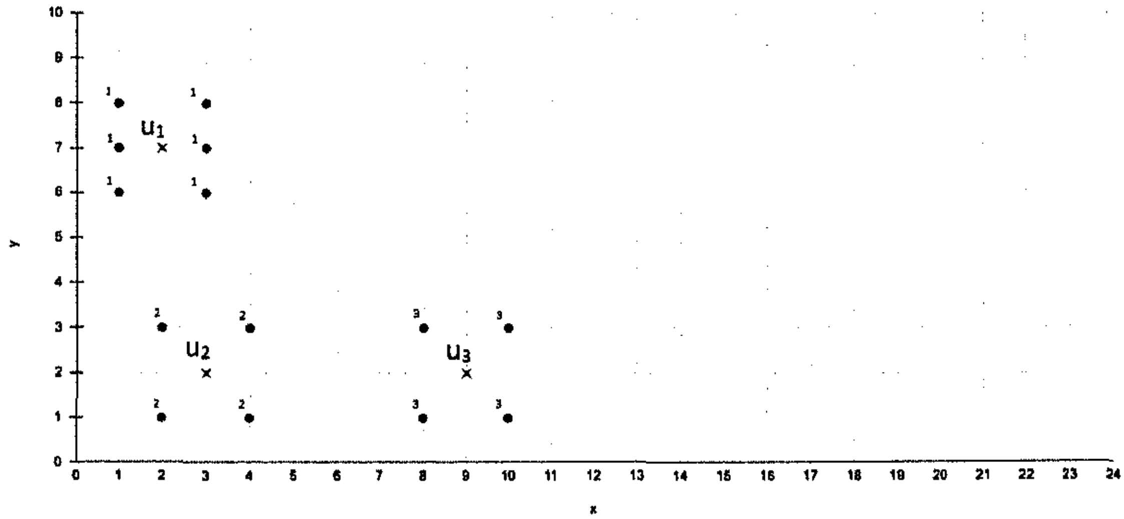
בהינתן תוצאת משחק (תוצאות הטלה של שתי קוביות), אנו רוצים לסווג באיזה קזינו המשחק התרחש. הסתברות ה-Prior של Casino Golden Peacock היא 0.6.

2. (5 נק') נתונה תוצאת המשחק הבאה: בהטלת הקוביה הראשונה התקבל הערך 1 ובהטלת הקוביה השנייה התקבל הערך 6. באיזה קזינו התרחש המשחק באמצעות מסווג מסוג Full Bayes? הראו את חישוביכם.
3. (5 נק') בהינתן אותה תוצאת משחק שהתקבלה בחלק (2), באיזה קזינו התרחש המשחק באמצעות מסווג מסוג Naïve Bayes? הראו את חישוביכם.
4. (5 נק') מהו ערך ה-Prior המינימלי הנדרש עבור Silver Turkey על מנת שמסווג Full Bayes יבחר ב-Silver Turkey בלי קשר לתוצאת המשחק? הסבירו \ הראו את חישוביכם.
5. (5 נק') באפשרותכם לשנות שני ערכים בטבלת ההסתברות המשותפת של קזינו Golden Peacock. בהינתן אותה תוצאת משחק שהתקבלה בחלק (2), משמע (1,6), שנו שני ערכים כך שמסווג Full Bayes יבחר בקזינו Golden Peacock תחת ערך הסתברות ה-Prior שקיבלתם בחלק (4). ההתפלגות החדשה שיצרתם צריכה להיות התפלגות תקינה.

שאלה 5 – k-Means (25 נקודות)

הערה: בכל הסעיפים של שאלה זו יש להניח מרחק אוקלידי.

1. בגרף הבא ישנן 14 נקודות ב-$\mathbb{R}^2$ וכן את המצב הנוכחי של הרצת אלגוריתם k-Means עם $k = 3$:



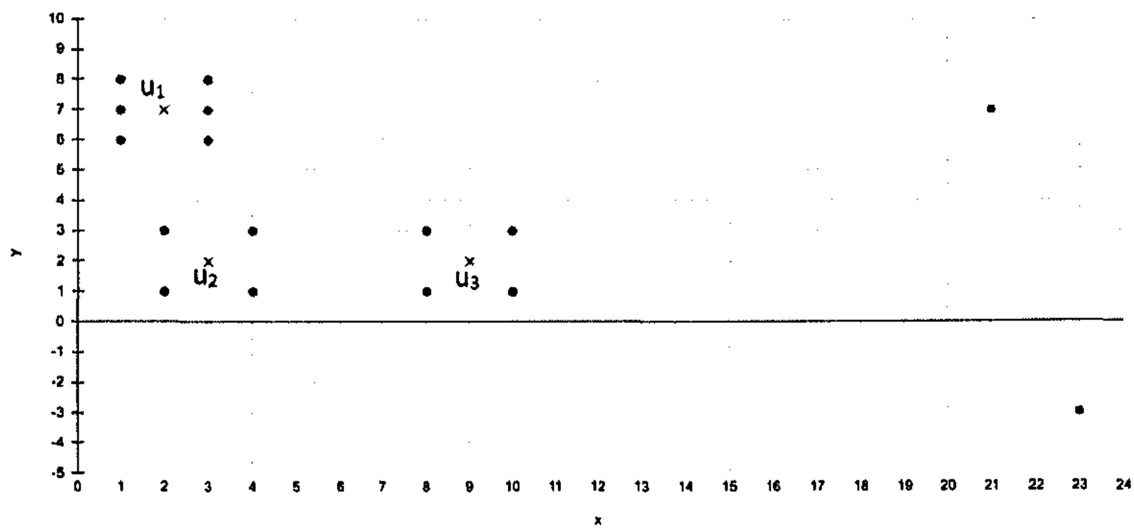
המספרים ליד כל נקודה מסמנים את הקלסטר אליה הנקודה משוייכת באיטרציה הנוכחית וסימוני ה-X מייצגים את המרכזים באיטרציה הנוכחית - u_1, u_2, u_3.

a) (3 נק') מהי הנוסחה לפונקציית ההפסד (Loss Function) של אלגוריתם k-Means?

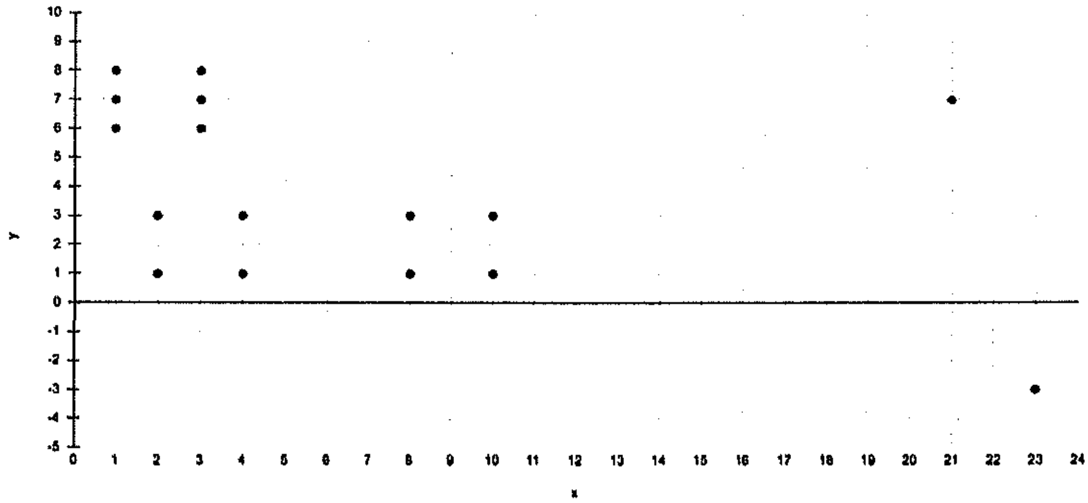
b) (3 נק') האם המצב המתואר בגרף הינו מצב יציב עבור ריצת האלגוריתם? האם הרצה של איטרציה נוספת תוביל לתוצאה שונה? הסבירו את תשובתכם.

לאחר הרצת האלגוריתם, התגלו 2 דוגמאות שהיו חסרות בדאטה המקורי. במקום להריץ מחדש את האלגוריתם (עם 16 נקודות במקום 14), החלטת להוסיף את הנקודות הללו ואז לבצע איטרציה נוספת מאותה נקודה אליה הגעת מקודם, החל משלב ההשמה לקלסטרים (בגרף לעיל). הרצת איטרציה נוספת של k-Means הביאה להשמה חדשה של נקודות לקלסטרים וכן למרכזים חדשים.

c) (2 נק') לאיזה קלסטרים ישויכו הנקודות החדשות? על גבי האיור הבא, ציין ליד כל נקודה חדשה לאיזה קלסטר היא תשויך.



d) (8 נק') על גבי האיור הבא, ציין את המרכזים החדשים באמצעות סימון X וכמו כן את ההשמות החדשות של כל הנקודות למרכז הקלסטר הקרוב אליהן ביותר (באופן דומה לאיור הראשון).



2. פונקציית ההפסד בסעיף (a) נקראת גם Inertia. עבור דאטה סט כלשהו המכיל 16 נקודות, הרצנו את אלגוריתם k-Means 16 פעמים עם ערכים משתנים של k החל מ-1 ועד 16. עבור כל הרצה, שמרנו את הערך ה-Inertia המתקבל עבור כל k.

a) (2 נק') איזו עמודה מבין העמודות A – F בטבלה להלן יכולה לייצג את התוצאות שקיבלת? (יש להניח כי עבור כל ערך של k האלגוריתם התכנס למינימום הגלובלי)

k	A Inertia	B Inertia	C Inertia	D Inertia	E Inertia	F Inertia
1	847	847	847	847	847	847
2	290	535	535	290	535	290
3	140	377	377	140	377	140
4	78	180	180	78	180	78
5	26	110	110	26	110	26
6	20	75	75	20	75	20
7	16	20	20	16	20	16
8	12	16	16	12	16	16
9	10	10	10	10	10	10
10	8	8	8	8	8	8
11	6	6	6	6	6	10
12	4	4	4	4	4	4
13	2.5	2.5	2.5	2.5	2.5	2.5
14	1	-1	0.5	1	1	1
15	0.5	-0.1	0	0.5	0.5	0.5
16	0	0	0	0.1	0	0

b) (5 נק') עבור כל אחת מהעמודות שלא בחרת, ציין מדוע היא לא יכולה להיות תוצאה של התהליך שהוגדר.

c) (2 נק') בהתאם לשיטת ה-"מרפק\ברך" שנלמדה בכתה למציאת ה-k האופטימאלי, עם איזה k הייתם בוחר לעבוד? נמק.

בהצלחה!

Standard formula sheet – IDC TASHPA

1. Distributions:

Normal $f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Binomial - $B(n,p)$ $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

Poisson $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

Geometric $P(X=k) = (1-p)^{k-1}p$

2. Decision Trees:

Gini $Gini(S) = 1 - \sum_{i=1}^{c}\left(\frac{|S_i|}{|S|}\right)^2$

Entropy $Entropy(S) = -\sum_{i=1}^{c}\frac{|S_i|}{|S|}\log\frac{|S_i|}{|S|}$

3. Gradient descent and update steps:

Linear regression $\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{d\in D}\left(h_\theta\left(x^{(d)}\right) - y^{(d)}\right)\cdot x_j^{(d)}$

Perceptron $w_j := w_j - \eta\sum_{d\in D}(o^{(d)} - t^{(d)})x_j^{(d)}$

Dual perceptron If $o^{(d)}\cdot t^{(d)} < 0$ then:

$\alpha_j = \alpha_j + \eta$

4. Logistic regression:

$P(h(x)=1) = \frac{1}{1+e^{-w^T x}}$

5. SVM:

Primal objective function $\frac{1}{2}\|w\|^2 + \gamma\sum_d \xi_d - \sum_d \alpha_d(t_d(w^T x_d + w_0) - 1 + \xi_d) - \sum_d \mu_d\xi_d$

s.t $\alpha_d \geq 0$ $\mu_d \geq 0$

Dual objective function $\sum_d \alpha_d - 1/2\sum_d\sum_e \alpha_d\alpha_e t_d t_e x_d^T x_e$

s.t $\sum_d \alpha_d t_d = 0,\ 0 \leq \alpha_d \leq \gamma$

6. EM (for Bernoulli distributions):

$New\, w_{A_j} = \frac{1}{N}\sum_{i=1}^{N} r(x_i, A_j)$

$p_{A_j} = \frac{1}{(New\, w_{A_j})N}\sum_{i=1}^{N} r(x_i, A_j)v(i)$

7. Linear Regression (closed form): $\theta^* = \underset{\theta}{\operatorname{argmin}}\|y - X\cdot\theta\|_2^2 = (X^T X)^{-1}X^T y$