

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

סמסטר ב' תשע"ו
Spring 2016

מבחן מועד א בלמידה ממוכנת
Machine Learning Exam A

Lecturer : Prof Ariel Shamir
Time limit : 3 hours
Additional material or calculators are not allowed in use!

Answer 5 out of 6 from the following questions (each one is 20 points)

מרצה: פרופ אריאל שמיר
משך המבחן: 3 שעות
אין להשתמש בחומר עזר ואין להשתמש במחשבוניס!

יש לענות על 5 מתוך 6 השאלות הבאות
לכל השאלות משקל שווה (20 נקודות)

בהצלחה!
Good Luck!

שאלה 1

- נתונים אוסף S של m דוגמאות x_i ($1 \leq i \leq m$) שלהן n תכונות (features) בהינתן k תתי-קבוצות S_i של S כך ש- $\bigcup_{i=1}^k S_i = S$ ו- $S_i \cap S_j = \emptyset$ $\forall i, j, 1 \leq i, j \leq k$.
א. הגדירי בנוסחה את המטריצות הבאות והסבירי מה משמעות האיברים שלהן (רמז: הפרידי בין האברים באלכסון ומחוצה לו):
1. Between-Class Scatter Matrix
2. Within-Class Scatter Matrix
ב. נניח כי μ_i הוא המרכז של קבוצה S_i (המרכז הוא הנקודה שערכה בכל מימד הוא הממוצע של הערכים של כל הנקודות בקבוצה S_i במימד זה) אנו מגדירים את הפונקצייה J באופן הבא $J = \sum_{i=1}^k \sum_{x \in S_i} |x - \mu_i|^2$ הסבירי מה J מודדת ומה הקשר (אם יש) בינה לבין המטריצות שהגדרת ב-א'.
ג. הסבירי בקיצור כיצד ניתן להשתמש באופטימיזציה של הפונקצייה J מסעיף ב' לבעיית קיבוץ (כאשר לא נתון מראש סיווג של האיברים ב- S) בשיטת k -means. מה המשתנים בפונקציה במקרה זה?
ד. הסבירי בקיצור כיצד ניתן להשתמש באופטימיזציה של פונקצייה J מסעיף ב' לבעיית סיווג (אם נתון מראש סיווג של האיברים ב- S) בשיטת LDA (Linear discriminant analysis). מה המשתנים בפונקציה במקרה זה?

שאלה 2

מס' דוגמא	מז"א	אטרקציה	סיווג
1	חם	שחייה	-
2	חם	ריקודים	+
3	חם	קזינו	+
4	חם	גולף	-
5	נוח	שחייה	-
6	נוח	קזינו	-
7	נוח	ריקודים	+
8	נוח	גולף	-
9	נוח	סקי	+
10	קר	סקי	+
11	קר	קזינו	-
12	קר	ריקודים	-

זוג גיקים מתכנן חופשה. הם החליטו להשתמש באלגוריתם Decision Tree כדי להחליט לאיזו חופשה לצאת בהתאם לשתי תכונות (features): מזג אויר שיכול לקבל 3 ערכים (חם, נוח, קר) וסוג האטרקציה בחופשה שיכולה להיות אחת מתוך (שחייה, ריקודים, קזינו, גולף, סקי). הם יצרו טבלה עם מס' דוגמאות, כאשר הסיווג הוא + אם הם אוהבים את החופשה ו- אם הם לא אוהבים אותה:

א. הגדירי מהו קריטריון Goodness of Split בבחירת ה-feature לפיצול הצומת בעץ החלטה והסבירי את הרעיון של קריטריון זה (מדוע הוא עוזר בבניית Decision Tree).

ב. הגדירי מהו Information-Gain (כולל נוסחה) והסבירי בכלליות איך הוא ממש את הקריטריון מסעיף א.

ג. הסבירי איזה חישובים צריך לעשות כדי למצא את התכונה (feature) שלפיה נפצל את שורש העץ בעזרת Information-Gain. אין צורך להגיע למספר סופי אלא רק להראות את החישובים שיש לעשות ולהסביר כיצד נחליט איזה תכונה נבחרה.

ד. הסבירי מדוע אלגוריתם המשתמש ב-Goodness of Split (Greedy)? הסבירי מה הבעיה באלגוריתם חמדן והציעי שיפור להתמודדות עם בעיה זו.

שאלה 3

נתונה קבוצת אימון המורכבת מ-m דוגמאות חלקן חיוביות (כלומר +1) וחלקן שליליות (כלומר -1). לכל דוגמא n תכונות (נניח כי הדוגמאות הן ב- R^n).

א. הגדירי מהי פונקציית החלטה (discriminate function) n מימדית וכיצד משתמשים בה בבעיית סיווג?

ב. הגדירי מהי פונקציית ההחלטה לסיווג שנוצרת באלגוריתם הפרספטרון על קבוצת האימון הנתונה (כולל נוסחה).

ג. הגדירי מהי פונקציית ההחלטה שנוצרת באלגוריתם LMS (least means square) המשתמש בשיטת gradient descent על קבוצת האימון הנתונה (כולל נוסחה).

ד. הגדירי מהו גרדיאנט של פונקציה n מימדית והסבירי איזה פונקציה מאופטמת בשיטת LMS (כולל נוסחה).

ה. נניח כי קבוצת האימון לא מסווגת רק לשתי מחלקות אפשריות, אלא ל-k מחלקות אפשריות כאשר $2 < k$. הציגי אלגוריתם שילמד מסווג ל-k מחלקות ומשתמש ב-LMS. הסבירי (כולל נוסחה!) מה תהיה פונקציית ההחלטה במקרה זה?

שאלה 4

נניח כי אנו רוצים לפתור בעיית סיווג של דוגמאות בין שתי מחלקות A ו-B.

א. הגדירי מהו מסווג Maximum A-Posteriori (MAP) ומאיזה בחינה הוא אופטימאלי.

ב. הגדירי מה ההבדל בין מסווג MAP למסווג מסוג Maximum Likelihood (ML), מתי ניתן להשתמש ב-ML?

ג. הגדירי מה היא בעיית רגרסיה בלמידה.

ד. הוכיחי כי ניתן להציג את פתרון בעיית הרגרסיה בעזרת מינימיזציה של סכום רבועי ההפרשים (sum of squared error) בצורה של היפותזה ML. בפרט הסבירי איזה הנחות עושים בהוכחה. רמז: הניחי כי קיים רעש (אי דיוקים) בנתונים ומדלי את הרעש כהאלי רנדומאלי.

שאלה 5

- א. נתון מרחב דוגמאות X המכיל מספר סופי של דוגמאות. הגדירי מהי דיכוטומיה על מרחב הדוגמאות.
- ב. בהנחה כי אנו מסווגים דיכוטומיות מעל X , הסבירי מהו מרחב היפוטזות ללא הטיה (unbiased) ומה גודלו?
- ג. הסבירי מהי דילמת bias-variance בהקשר של למידה חישובית.
- ד. הסבירי כיצד ניתן באלגוריתם K-Nearest Neighbors להשפיע על ה-bias מול ה-variance ומדוע.
- ה. הסבירי כיצד ניתן באלגוריתם decision tree להשפיע על ה-bias מול ה-variance ומדוע.

שאלה 6

נתונים שני חסמים שלמדנו על מספר הדגימות m בבעיית למידה:

- $m \geq \frac{1}{\varepsilon} (\ln |H| + \ln \frac{1}{\delta})$
- $m \geq \frac{1}{\varepsilon} \left(8vc(H) \log_2 \frac{13}{\varepsilon} + 4 \log_2 \frac{2}{\delta} \right)$

(א) הגדירי מהם $\delta, \varepsilon, |H|, vc(H)$

(ב) הסבירי מתי ניתן להשתמש בחסם הראשון ומתי בחסם השני. מדוע יש שניים?

(ג) בהינתן מרחב היפוטזות של מלבנים שבהם פנים המלבן מהווה את ההיפוטזה החיובית שהם מהצורה $(a \leq x \leq b) \wedge (c \leq y \leq d)$, כאשר a, b, c, d הם מס' שלמים בקטע $[0, n-1]$. הראי כי מספר המלבנים השונים הקיימים בתנאים האלה בתחום הסגור $(0,0) \times (n-1, n-1)$ הוא $\left(\frac{n(n+1)}{2}\right)^2$.

(ד) מהו מימד VC של מרחב ההיפוטזות מסעיף ג?

בסעיפים הבאים נניח כי אנו לומדים במרחב דו מימדי בעל ערכים בין 0 ל-99 בכל מימד.

- (ה) נתון אלגוריתם למידה המבטיח טעות אימון 0. כמה דוגמאות אימון צריך הלימוד במרחב ההיפוטזות מסעיף ג כדי להבטיח בהסתברות לפחות 95% היפותיזה עם טעות לכל היותר של 0.15? הראי את החישוב בלבד (אין צורך לקבל תוצאה סופית).
- (ו) נניח עכשיו שהערכים a, b, c, d הם מספרים ממשיים בתחום הנתון (ולא מספרים שלמים כמו שהוגדר מקודם). האם התוצאה מסעיף ה' תשתנה? אם לא, מדוע? אם כן, הסבירי כיצד תחושב כמות דוגמאות האימון הנדרשת? (אין צורך לקבל תוצאה סופית).

בהצלחה!

