עקרונות למידה חישובית:

ש: מה ההבדל בין קלסיפיקציה לבין רגרסיה?

ת: סיווג עבור ערך בדיד מתוך רשימה סגורה של אפשרויות, לבין חיזוי ערך רציף

ש: הסבירו את ההבדל בין למידה פרמטרית לבין למידה לא פרמטרית

ת: בלמידה פרמטרית המודל מיוצג ע"י אוסף מקדמים המגדירים אותו, בעוד שבלמידה לא פרמטרית אנו מייצרים גבולות החלטה בצורה מלאכותית, כך שמתקבל מרחב עבור סיווגים

ש: מבין המודלים הבאים – KNN, SVM, עץ החלטה (Decision Tree), רגרסיה לינארית ורגרסיה לוגיסטית, מיהו מודל בעל למידה פרמטרית, ומי בעל למידה לא פרמטרית?

ת: פרמטרית: SVM, רגרסיה לינארית, רגרסיה לוגיסטית. לא פרמטרית: KNN, Decision Tree

ש: מהו קידוד one-hot ומתי נשתמש בו? מה יהיה גודלו, וכתלות במה?

ת: ווקטור בינארי בגודל כמות הסט המאפיין של עמודה מסוימת, כאשר העמודה מייצגת משתנה נומינלי

רגרסיה:

ש: הוכח\הפרך – עבור כל משתנה מסביר X ומשתנה מוסבר Y, אם נשתמש ברגרסיה לינארית – נוכל לבצע אותה בהכרח גם עבור המצב בו Y משתנה מסביר ו-X משתנה מוסבר

ת: הוכחה – פיתוח מתמטי: Y = aX +b -> Y-b = aX -> X = (Y-b)\a

ש: באיזה מצב לא נוכל "להפוך" את ישר הרגרסיה הלינארית?

ת: כאשר מדובר בקו קבוע מהצורה y=k

ש: מה ההבדל ברמת המשתנה המוסבר בין רגרסיה לוגיסטית לבין רגרסיה לינארית?

ת: לוגיסטית עבור חיזוי משתנה בינארי, לינארית עבור משתנה רציף

ש: עבור רגרסיה לוגיסטית, כיצד נקבל החלטה עבור פלט בין 0 ל-1 של הרגרסיה?

ת: נגדיר threshold = 0.5, מה שמעליו יהיה 1, מה שמתחתיו יהיה 0

ש: עבור רגרסיה לינארית, למדנו שהפלט נראה מהצורה y = b0 + b1X. מה משמעות המקדם b0?

ת: מקדם חופשי, המייצג bias של המודל

ש: עבור רגרסיה לוגיסטית, למדנו שמשוואת המודל היא מהצורה: . ברמת הגרף הנוצר מעקומת הרגרסיה, איך משפיעים על ההצגה הגרפית הפרמטרים ?

ת: משפיע על מיקום הקו ה"אנכי", בעוד ש- משפיע על תלוליות העקומה.

קלסיפיקציה:

ש: א - מנה 3 חסרונות של אלגוריתם k-nearest neighbors

ת: זמן ריצה ארוך ככל שכמות הדוגמאות גדלה, רגישות לדוגמאות חריגות, אי התכנסות של חישוב מרחק כאשר כמות המימדים גדלה

ש: ב - מהו פיתרון אפשרי לכל חסרון שכזה?

ת: להשתמש בדיאגרמת וורונוי ולסלק נקודות שמוקפות באותו תחום החלטה, סילוק דוגמאות חריגות ע"י דיאגרמת וורונוי, משקול שונה לכל עמודה

ש: תארו 2 גישות שונות לפחות עבור cross validation והסבירו אותן

ת: leave one out, k-fold

ש: הסבר את ההבדל בין one-vs-one לבין one-vs-all במודל SVM, ותאר את כל אחת משתי הגישות הנ"ל. באופן כללי – עבור n מחלקות, כמה מודלים **בדיוק** ייבנו בשיטת one-vs-one, וכמה בשיטת one-vs-all?

ת: עקרון מג'וריטי מול רוב בפעם אחת (OVO – n\*(n-1)/2 מודלים, OVA – n מודלים)

ש: מה משמעות פרמטר ה-C במודל SVM?

ת: "עונש" על נקודות שהן support vector המייצרות overfit למודל

ש: איזה סוג למידה היא Instance-based learning? תן דוגמא למודל המשתמש בה, וכיצד זה בא לידי ביטוי?

ת: KNN כדוגמא, כאשר אנו ממש שומרים פיזית את כל בסיס הנתונים כדי לבצע פרדיקציה

ש: הגדר והסבר מה הן נקודות Support Vector? איזה עיקרון הן באות ליישם ברמת בניית מודל ה-SVM והמרחק שלו מהן?

ת: נקודות ה"תמיכה" של מודל SVM השומרות על עיקרון המקסימום שוליים

ש: מודל SVM איננו מותאם מלכתחילה לבעיות multi-class (ריבוי מחלקות), אלא לבעיות סיווג בינאריות. תאר והסבר שתי שיטות לבניית מודל SVM עבור בעיות קלסיפיקציה multi-class?

ת: one-vs-one, one-vs-all

ש: איזה פרמטר במודל SVM מאפשר שליטה בבניית המודל המפריד, ע"פ ה-support vector? ומהו ה-tradeoff בין ערך גדול מדי, לערך קטן מדי?

ת: פרמטר C, שהוא עלות (COST) וה-tradeoff הוא overfit מול underfit

Clustering:

ש: באיזו דרך ניתן לחשב K אופטימלי (כמות הקלאסטרים) עבור אלגוריתם k-means? הסבר כיצד היא באה לידי ביטוי בצורה גרפית, וכיצד תבחר את K?

ת: ה-inertia

ש: תאר את אלגוריתם K-means

ת: תיאור האלגוריתם

ש: מהו תנאי העצירה באלגוריתם k-means?

ת: איטרציה i+1, שעבורה אין שינוי במיפוי הנקודות לקלאסטרים ביחס לאיטרציה i

ש: תן דוגמא לאלגוריתם clustering אשר רגיש ל-outliers (דוגמאות חריגות) והסבר מדוע? כיצד תתגבר על מכשול זה?

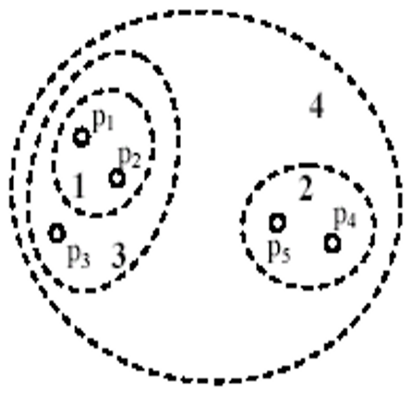
ת: k-means, לאור ה-centroids המחושבים. ניתן למחוק דוגמאות שרחוקות מאוד מה-centroids ע"י הגדרת threshold, או לבצע sampling אקראי

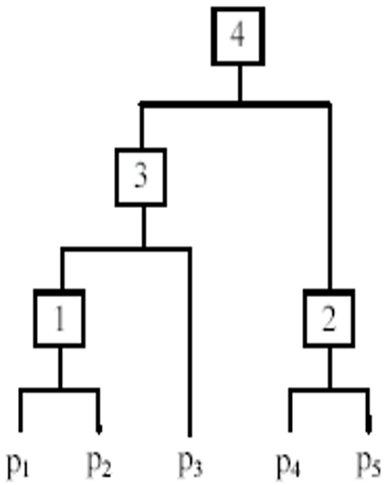
ש: מהן שתי הנחות היסוד שאלגוריתם k-means מניח ביחס למבנה של הקלאסטרים?

ת: היפר-אליפטיות, היפר ספריות

ש: הסבר ותאר מהו דנדרוגרם? בנוסף, הסבר את שיטות bottom-up ו-top-down לבנייתו.

ת: רצף קלאסטרים מקונן, בצורת "עץ". ה-bottom-up מניח שכמות הקלאסטרים היא כמות הדוגמאות ומשם בונה עד השורש, וההפך עבור top-down כאשר כמות הקלאסטרים בהתחלה היא 1

ש:  צייר את הדנדרוגרם עבור הגרף הבא:

ת: 

ש: למדנו 3 שיטות עבור חישוב מרחק בין שני קלאסטרים. מהן? הסבר אותן בפירוט

ת: single link, complete link, average link

הסתברות, מדדי שגיאה ומטריקות:

**שאלה**:

נאמר שהדאטה שלי מורכבת מ-150 דגימות והחלוקה היא: 100 דגימות של "כלב" ו-50 דגימות של "חתול". בניתי מודל סיווג שידע לתייג "כלב" או "חתול".

המודל שלי תייג 90% מהכלבים כ-"כלב" ו-40 אחוז מהחתולים כ-"חתול".

בנה את ה-confusion matrix.

**תשובה**:

|  |  |
| --- | --- |
|  |  |
|  |  |

**שאלה**:

יש 4 בדיקות למחלת השפעת:

* ההסתברות שהרופא יבדוק אותך בבדיקה A היא 50%, והיא צודקת ב- 95%.
* ההסתברות שהרופא יבדוק אותך בבדיקה B היא 30% והיא צודקת ב- 70%.
* ההסתברות שהרופא יבדוק אותך בבדיקה C היא 15% והיא צודקת ב- 60%.
* ההסתברות שהרופא יבדוק אותך בבדיקה D היא 5% והיא צודקת ב-55%.

**סעיף א'**: אדם נבחר באופן אקראי, הלך לרופא, נבדק, מה ההסתברות שתוצאות הבדיקה שלו נכונה?

**פתרון א':**

**סעיף ב'**: נאמר שהבדיקה של אותו אדם הראתה תוצאה שגויה. מה ההסתברות שהוא נבדק בבדיקה C?

**פתרון:**

**שאלה**:

יש בכד 4 כדורים כחולים, 2 כדורים אדומים ו-3 כדורים צהובים.

מוציאים כדור באופן אקראי מבלי להחזיר אותו לכד.

מה ההסתברות ששלפנו בפעם הראשונה כדור צהוב בהינתן זה שבפעם השנייה לא שלפנו כדור צהוב?

**פתרון:**

**סעיף ב'**: נאמר שהבדיקה של אותו אדם הראתה תוצאה שגויה. מה ההסתברות שהוא נבדק בבדיקה C?

**פתרון:**

**שאלה**:

עבור המטריצה הבאה, חשב: Accuracy, Precision, Recall (אין צורך במספר מדויק, מספיק לרשום בצורת שבר מסוג ).



**תשובה**:

שאלות קוד:

**שאלה**:

נתון קוד.

מצא את שתי הבעיות בקוד וכתוב כיצד יש לתקן אותן.

**תשובה**:

בשורה 2 נחליף את ב- כדי שהמודל ימצא התאמה בין הדגימות לבין התוויות של ה-train. בשורה 4 נחליף את ב- כיוון שהחיזוי שלנו התבצע על הדגימות ב-test ולכן נרצה לבדוק את נכונות המודל שלנו ביחס לתוויות של ה-train.