**פרק 6 : מודלים קלאסיים נוספים**

חלק תיאורטי:

1. SVM

Read about SVM from ISLR pages 337-359

Answer questions 3,6 in exercise section

Then proceed with the next question:

In the next exercises assume the data is 2 dimensional and distributes uniformly on [0,1] X [0,1] unless otherwise stated.

**Part 1**

In this exercise we assume the hypothesis we want to learn is classification by y=1-x:

Class = 1 if y>1-x, otherwise 0

Create a train dataset by generating 1000 labeled points

Create a test dataset by generating 1000 unlabeled points

Classify each point according to a different kernel

Show them on the same plot (one for every kernel). Correctly classified points in blue, misclassified in red. Try at least 3 different kernels which you see fit. Also plot the separating line.

What is the difference in total train and test time between different kernels?

Choose the appropriate kernel and use it in the next parts:

**Part 2**

Extend the hypothesis for higher dimension while keeping the training and test sets balanced. Classify each point according to your chosen kernel. How are the training and test times depend upon the dimension? (what is the relation)

**Part 3**

Denote the number of train samples N, and the dimension P. Test samples number is fixed. For every region N<<P, N=P, N>>P show the test error. Plot all results one informative graph.

**Part 4**

Repeat the exercise 1 with hypothesis of circle locates at (0.5,0.5) with radius of size 0.25. No need to plot separating line.

Try this using linear and using polynomial kernel of deg 2.

1. Bayesian:

**תיאורטי:**

* לקרוא את פרק 5.6 - בספר deep learning בתור התחלה.
* לצורך המחשה בגרפים יש לקרוא את פרקים 3.3-3.3.1 בספר Bishop – Pattern Recognition

**הרחבת ביניים תיאורטית (למעוניינים).**

* **להרחבה** והסבר יותר מפורט על הבדלי הגישות המהותי ניתן לקרוא את השיעור מMIT אשר מסביר לעומק את היתרונות והחסרונות של שתי הגישות השונות של סטטיסטיקה להבנת התפלגויות וחיזוי.
* כמו כן ניתן לקרוא על כל השיטה עם דוגמאות מהספר המקורי שמלווה את הקורס בMIT:

קרא מתוך הספר introduction to probability את פרק 8 - סטטיסטיקה בייסיאנית.

**מעשי:**

שיטות בייסיאניות נדירות היום במרבית בעיות הML שתפגוש, אך יש לה יתרון מאוד גדול כאשר יש לך prior ברור מתוך ידע עסקי, ומעט דאטא, דבר שאפשרי להיתקל בו במסגת העבודה שלנו.

* בבלוגפוסט הבא הכותב מתאר את השימוש שלו בספריה שמממשת BLR ועושה השוואה של הצלחת מודלים שונים: הוא מנסה לחזות ציונים של סטודנטים על סמך נתונים גולמיים שלהם (יחסית מעט סטודנטים, עם הרבה נתונים), ובנוסף הוא מראה כיצד ניתן להסביר את המודל.

הוא טיפה חופר, אפשר לקרוא ברפרוף, אבל לשים לב לדגשים הבאים:

* איך הוא בונה את המודל
* השימוש בMCMC לדגימת מרחב הפרמטרים
* איך הוא משתמש במודל למתן פרדיקציה והערכת מדד ביטחון.

[https://towardsdatascience.com/bayesian-linear-regression-in-python-u](https://towardsdatascience.com/bayesian-linear-regression-in-python-using-machine-learning-to-predict-student-grades-part-2-b72059a8ac7e)

[sing-machine-learning-to-predict-student-grades-part-2-b72059a8ac7e](https://towardsdatascience.com/bayesian-linear-regression-in-python-using-machine-learning-to-predict-student-grades-part-2-b72059a8ac7e)

להסבר נוסף ויותר אינטואיטיבי על MCMC ניתן לראות את ההסבר מתוך קטע הבלוג הבא, אשר מפרט יותר טוב על נקודה זו:

<https://towardsdatascience.com/an-introduction-to-bayesian-inference-e6186cfc87bc>

רוב הדוגמאות שראיתם כאן הם על רגרסיה, כמובן שניתן להפוך כל מודל למודל בייסיאני, אפילו רשתות נוירונים - תחום מחקר אשר מנסה להשלים את הפער עם ההצלחות האחרונות של רשתות נוירונים לא בייסיאניות (תחום עתיק בפני עצמו) אשר נבנו מתוך ניסוי וטעיה אמפירים בניגוד לרשתות בייסיאניות.