**תרגול 1**

חזרה על סטטיסטיקה

שונות- הפיזור של ההתפלגות, כמה התצפיות קורבות לממוצע

התפלגות נורמלית- מורכבת ממוצע ושונות

רגרסיה לינארית- משתמשים במשוואה של קו ישר, על מנת למדל את הקשר בין x לy.

מסמנים נקודה בyi, נסמן את המרחק של כל הנקודות yi לקו הממוצע של הנקודות yi .

ניקח את סכום המרחקים(השאריות-residual) , ככל שסכום ריבועי השאריות קטן יותר כך הפונקציה מתארת את הנתונים בצורה טובה יותר.

שלבים של למידת מכונה

1. איסוף נתונים
2. בדיקת המידע וויזואליזציה שלו
3. עיבוד מוקדם של המידע
4. בניית מודל
5. הערכה של המודל

סוגי למידת מכונה

Supervised learning- משהו שמנסים לחזות (מה המחיר של בית, האם זה כלב או חתול), יש משתנה-label שמנסים לחזות אותו, המשתנים שאנו משתמשים בהם כדי לחזות את המשתנה הם features, משמש לחיזוי

לדוגמה:

מקבלים נתונים של תפוחים וקאפקייקס, מכניסים את הנתונים למודל של למידת מכונה, המודל לומד מה זה תפוח ומזה קאפקייקס, לאחר מכן משתמשים בנתונים שלא היו באימון ומודל מוצלח יצליח להבין האם הנתונים מהפרט שהתקבל הוא תפוח או קאפקייק.

רגרסיה- מנסים לחזות משתנה מספרי

סיווג- מנסים לחזות חלוקה לקטגוריות

Unsupervised learning- מנסים להבין דברים על הנתונים, לדוגמה clustering- חלוקה של הנתונים לקבוצה, density estimation- לחזות איך הנתונים מתפלגים, צמצום ממדים.

לדוגמה:

מקבלים נתונים של תפוחים וקאפייקס, מכניסים לאימון על מודל של למידת מכונה ומגדירים שיחלק את הנתונים ל2 קבוצות. אין labels, הוא מחלק את הנתונים בצורה שנראית הגיונית, לא מנסים לחזות משהו, משמש לאנליזה.

Reinforcement learning- למידה מטעויות

על מנת להעריך כמה מוצלח המודל, מחלקים את המודל לtraining data ו test data -בדרך כלל חלוקה של 80% אימון 20% מבחן

שיטה נוספת היא חלוקה נוספת לvalidation data.

היפר-פרמטרים: פרמטרים שמשתמשים בהם כדי להגדיר את האלגוריתם, לדוגמה להחליט את עומק עץ ההחלטה (כמה רמות), על מנת לשפר את הביצועים של המודל.

קרוס ולידציה k fold- בונים 5 עצי החלטה שונים באמצעות שימוש בחלקים שונים מהנתונים על מנת להעריך כמה טובים הביצועים של כל סוג עץ.

במטריצה כל שורה זה עץ, המלבנים הכחולים בכל שורה זו הולדיציה- ה20% מהנתונים שהם המבחן של העץ.

קרוס ולידציה של leave one out, כל פעם להישאר חלק אחר מהנתונים מחוץ לאימון ולבסוף לבחון עליו.