

פרוייקט קורס סופי

קורס הדמיית נתונים סמסטר ב' 2021

שם: כפיר גולדפרב

אימייל: kfir.golfarb@msmail.ariel.ac.il

סיכום פרויקט:

חלק א':

בסמסטר שעבר ביצעתי בעיית קלאסיפיקציה על דאטא של תלמידים עם 3 ציונים שונים, לקחתי ממוצע של שלושת הציונים עבור כל תלמיד, מי שעבר את ה-60 נחשב עובר, ומי שלא נחשב נכשל, ניסיתי לבצע קלאסיפיקציה בינארית בין עובר לבין נכשל לפי הפרמטרים הנתונים בדאטא (הסברתי היטב במחברת איזה פרמטרים יש ומה כל אחד אומר), ביצעתי קלאסיפיקציה בעזרת שלושה מודלים:

KNN, Logistic Regression, Decision Tree

הדיוק המקסימלי היה של ה-Logistic Regression עם דיוק של 77.60%,

בסמסטר זה לקחתי את אותה מחברת וביצעתי עליה ensemble learning עם המודלים:

soft and hard Voting Classifier (על שלושת המודלים מסמסטר קודם), Bagging, Pasting, Random Forests, AdaBoost, Gradient Boosting, Extreme Gradient Boosting,

המודל בעל הדיוק הגבוה ביותר הוא adaboost עם דיוק של 78%.

חלק ב':

בחלק ב' ניסיתי למצוא מודל קלאסיפיקציה הכי טוב עבור Fashion MNIST Dataset, לאחר הורדת פיצ'רים בעזרת PCA, מספר הפיצ'רים שבחנתי הם: 24, 50, 154.

המודלים שבהם השתמשתי בכל אחד מהם:

K-Means, Bagging, Pasting, Random Forest, AdaBoost, Extreme Gradient Boosting, Grid Search with Random Forest, Grid Search with Extreme Gradient Boosting.

המודל עם הדיוק הכי גבוה שמצאתי היה Extreme Gradient Boosting עם 154 פיצ'רים בדיוק של 88.14%.

חלק ג':

בחלק 3 ניסיתי לבצע בעיית קלאסיפיקציה בין חתול לכלב עבור Cat & Dog Dataset, לאחר הורדת פיצ'רים בעזרת PCA, מספר הפיצ'רים שבחנתי הם:

1. 335 (תמונות 100*100 פיקסלים greyscale),
2. 300 - (תמונות 30*30 פיקסלים no greyscale),
3. 185 - (תמונות 30*30 פיקסלים no greyscale).

המודלים שבהם השתמשתי בכל אחד מהם:

Bagging, Pating, Random Forest, AdaBoost, Extreme Gradient Boosting, Grid Search with Random Forest, Grid Seach with Extreme Gradient Boosting

המודל עם הדיוק הכי גבוה שמצאתי היה Grid Seach with Extreme Gradient Boosting עם 300 פיצ'רים בדיוק של 66.20%.

חלק ד':

בחלק ד' ניסיתי למצוא מודל קלאסיפיקציה הכי טוב עבור Hands Dataset, המודלים שבהם השתמשתי הם:
KNN, Decision Tree, Logistic Regression, Bagging, Pasting, Random Forest, AdaBoost, Stacking,
soft Voting, hard Voting, Extreme Gradient Boosting, Grid Search with Random Forest, Grid Search
with Extreme Gradient Boosting,

המודל עם הדיוק הכי גבוה שמצאתי היה KNN בדיוק של 88.69 %.