Module 9: Supervised Learning- 2

Case Study

Case Study

Objective:

- Employ SVM from scikit learn for binary classification.
- Impact of preprocessing data and hyper paramter search using grid search.

Questions:

1. Load the data from "college.csv" that has attributes collected about private and public colleges for a particular year. We will try to predict the private/public status of the college from other attributes.

2. Use LabelEncoder to encode the target variable in to numerical form and split the data such that 20% of the data is set aside for testing.

3. Fit a linear svm from scikit learn and observe the accuracy.
   [Hint: Use Linear SVC]

4. Preprocess the data using StandardScalar and fit the same model again and observe the change in accuracy.
   [Hint: Refer to scikitlearn's preprocessing methods]

5. Use scikit learn's gridsearch to select the best hyperparameter for a non-linear SVM, identify the model with best score and its parameters.
   [Hint: Refer to model_selection module of Scikit learn]