- Training Classification models

1 Introduction

Once your data has been processed, it is time to use it to train models. We will start with classification models. This homework is intended to give you a brief introduction on how to train classification models. It covers the following topics:

- Perform classification with:

  Decision Tree Classifier.

  K-Nearest Neighbors Classifier.

  Logistic regression Classifier.

- Generate predictions from these models.
- Plot confusion matrices.
- Find accuracies of these models

2 Tasks

Complete the following tasks, and export your work as a pdf file. Submit the report along with the additional files (saved figures and data files) on Canvas.

1. Load the `patient.csv` data file.
2. Create a DataFrame called *data* using the following variables:

   Age

   Gender

   Systolic

   Diastolic

   Smoker

   Set Smoker as the response variable (Y), and select everything else as a predictor variable (X).

3. Train a Decision Tree Classifier Model

   Create a decision tree classifier.

4. Plot the Systolic vs Diastolic data. Color by their response variable. And change the style based on whether the prediction is correct or not.

   Are the clusters clearly identifiable?

   Change the X-axis to other parameters such as gender and age. Do they have better clustering?

   What is the Accuracy for the decision tree model? Is it a good?

5. Plot the Confusion Matrix.

   How many false positives and false negatives do you have?

6. Repeat the above steps, but choose only the Systolic and Diastolic values as predictor variables.

   Does this improve the accuracy of the model?

7. Make predictions using the trained model for the following inputs:

   Diastolic=90 mmHg, Systolic=125 mmHg

   Diastolic=85 mmHg, Systolic=130 mmHg

8. Repeat the steps above for a KNN Model.

9. Repeat the steps above for a Logistic Regression Model.

   What are the probability values generated in the Logistic regression model using the input data.

10. What was the best and worst model? What are their Accuracies?