Lab1

1. Introduction

In this lab, I analyzed 3 datasets: the iris dataset, the mushroom dataset, and the heart dataset. All these datasets are available from Kaggle.com.

1. Objectives

The objective was to analyze the datasets and fit appropriately accurate models.

1. Approaches/Methods

In this lab, I used several different methods of supervised and unsupervised learning. I used multiple linear regression, support vector machines, k-means clustering, k-nearest neighbors, and naïve Bayes.

1. Datasets

Here are listed the datasets that were used in my lab:

https://www.kaggle.com/uciml/mushroom-classification

https://www.kaggle.com/uciml/iris

<https://www.kaggle.com/ronitf/heart-disease-uci/downloads/heart-disease-uci.zip/1>

1. Conclusion

The datasets were all relatively clean, so getting good results from the required models was straightforward. I achieved a good R^2 score (0.73) for the multiple linear regression on the mushrooms data set. The naïve Bayes, support vector machines, and k-nearest neighbors all achieved accuracy scores of over 95% on the iris dataset, which is very good. And the silhouette score of the heart dataset, an unsupervised learning attempt, was 0.12, which isn’t exactly great, but not terrible, either. Overall, these are acceptable results.