Lab 5

Goal: Create a logistic regression classifier based on different selected features.

1. Fit a logistic regression to the modified iris dataset, with the species as the label. Train using the first 100 samples and test using the last 50 samples.
    a. Modified Iris Dataset can be downloaded from the course website or by running the first code block provided in the Colab notebooks below
2. Run through the feature selection process, using either SelectKBest or VarianceThreshold to find the best 2 features and store this new cleaned dataset.
3. Fit another logistic regression to the cleaned dataset and compare the performance. Once again use the first 100 samples to train and last 50 to test.
4. **Bonus Task**: Use the wine dataset. Try different combinations of the hyperparameters (for example, n_estimators and max_depth on random forest) and number of features (SelectKBest) and try to get as high an accuracy as possible. The answer key uses random forest. You are welcome to choose a different classifier.

LAB 5S -- Overfitting and Feature Selection

Scaffolded lab template

LAB 5B -- Overfitting and Feature Selection

Blank lab template