CSE 40625 — Machine Learning
Assignment 3 (10 points)
Due Date: Mar. 2, 2017 (Sakai Drop Box)

Gaussian Naïve Bayes

Overview

The purpose of this assignment is for you to implement Gaussian naïve Bayes, a simple but powerful machine learning model. Gaussian naïve Bayes is a supervised classifier that makes predictions by applying Bayes Theorem with the "naïve" assumption of feature independence, and further assumes the likelihood of features to be Gaussian. Materials for the assignment, including the dataset, expected output, and template code can be found here on GitHub.

You may use Python libraries for handling data preprocessing and visualization, including but not limited to NumPy, SciPy, pandas, and Matplotlib, but *you may NOT use any Python libraries that employ machine learning models, including but not limited to scikit-learn, StatsModels, TensorFlow, or Orange*. Your solution to the assignment should be individually submitted.

Dataset

You will use the "digits" dataset on handwritten digit classification with all 10 classes (labeled from 0 to 9) for this assignment. The data is provided in comma-separated (CSV) file format. For all rows, the last column designates the class (y) and the remaining columns designate features (X). The first row consists of the feature and class names.

Procedure

Use the digits dataset discussed above as the input to a Gaussian naïve Bayes model. Empirically estimate the mean (theta) and variance (sigma) of each feature conditioned on each class value. Use the thetas and sigmas to calculate the posterior log likelihood for each instance under the Gaussian assumption. To avoid the evaluation of log(0.0), add an offset of 10e-5 to each calculated sigma. Predict each target value that maximizes the calculated (log) likelihood.

Output

Your code should output the accuracy of the prediction results when predicting over the same data used to fit the model, a blank line, and the confusion matrix of the same results.

Example output:

```
0.864

Predicted   0    1  2  3  4  5  6  7  8  9
Actual
0         548    2  0  0  0  1  1  0  2  0
…
```

Submission

Please submit a Python executable (gaussiannb.py) file of your code to your Sakai Drop Box. *Should you run into any problems, please feel free to email or meet with the instructor.*