

CS-584 – Assignment 2 (5%)

Generative learning

Due by: March 10, 2016

Assignment Specifications

In this assignment you will implement algorithms for generative learning. You have to use two or more external data sets one with continuous features and one with discrete features. Links to data sets are available on the course web-page (e.g. the UCI Machine Learning Repository <http://archive.ics.uci.edu/ml/>). It is essential that you evaluate the performance of each algorithm you implement and the effects of varying different parameters on the performance of the learning algorithm. Use cross validation to test performance. The assignment should be implemented using Python. You may use all Python functions except for the one you are asked to implement. You may still use the Python function you are asked to implement as a mean for verifying the accuracy of your implementation. The grade for this assignment will be based in part on the performance of your implementation and on the thoroughness of your evaluation.

1. 1D 2-class Gaussian discriminant analysis

- (a) Select a 2-class dataset with continuous 1D features.
- (b) Estimate the model parameters and compute a discriminant function based on the distribution in each class.
- (c) Classify the examples and measure your error. Compute the confusion matrix, precision, recall, F-measure, and accuracy.

2. nD 2-class Gaussian discriminant analysis

- (a) Select a 2-class dataset with continuous nD features.
- (b) Estimate the model parameters and compute a discriminant function based on the distribution in each class.
- (c) Classify the examples and measure your error. Compute the confusion matrix, precision, recall, F-measure, and accuracy.
- (d) Plot the precision-recall curve and measure the area under curve.

3. nD k-class Gaussian discriminant analysis

- (a) Select a k-class dataset with continuous nD features.
- (b) Estimate the model parameters and compute a discriminant function based on the distribution in each class.
- (c) Classify the examples and measure your error. Compute the confusion matrix, precision, recall, F-measure, and accuracy.

4. Naive Bayes with Bernoulli features

- (a) Select a 2-class dataset with binary nD features. This dataset needs to be derived from text documents.
- (b) Estimate the model parameters and compute a discriminant function based on the distribution in each class. Make sure to use the Naive Bayes assumption.

- (c) Classify the examples and measure your error. Compute the confusion matrix, precision, recall, F-measure, and accuracy.

5. Naive Bayes with Binomial features

- (a) Using maximum likelihood derive the parameter estimate equations for Naive Bayes with Binomial features: write the log likelihood function, compute its derivative, equate the derivative to zero, and solve for the parameters. This is a written step that does not require programming.
- (b) Select a 2-class dataset with discrete nD features. This dataset needs to be derived from text documents.
- (c) Estimate the model parameters and compute a discriminant function based on the distribution in each class. Make sure to use the Naive Bayes assumption.
- (d) Classify the examples and measure your error. Compute the confusion matrix, precision, recall, F-measure, and accuracy.

Follow the submission instruction of assignment 1.