

Project 4

Due Dec 9th 2012

In this project you will implement two kinds of classifiers – (1) Decision Tree and (1) Naïve Bayes.

1. The task here is to build a Decision Tree Classifier for predicting credit worthiness of applicants. The data set for building and testing the classifier is here: <http://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening>. The data set has 3 types of attribute values – categorical, integers and reals. Some attribute values are also missing. Your implementation must incorporate techniques for addressing overfitting. The project report must include a detailed analysis of your implementation including how well it handles overfitting, its precision and any other performance metrics you think is important.
2. The task here is to build a Naïve Bayes Classifier for classifying handwritten digits. The file Data.zip consists of labeled training and test set. Use the training set for building the classifier. The feature set consists of a single binary indicator feature value for each (i,j) pixel. Its value is 1 if the pixel is foreground and 0 if it is background. Your implementation should handle zero counts. The project report must include quantitative performance measurements, in particular the precision for each digit