Study and Implementation ofClassifiers

**1. Abstract**

**The goal of this project** is to develop a pattern recognition system that operates on a  
given real-world dataset. In this work we develop various classifiers to work  
on a real-world data set chosen and classify the samples into the corresponding classes. In  
this project I have chosen the “Wireless Indoor Localization” dataset to work on. I have  
implemented a 3 class classification problem. Through the project I have implemented  
various classifiers like Navie Bayer, SVM, Perceptron, Random Classifier and the support vector machine and performed a comparative study of their performance. The dataset was first processed so that the classifiers can be trained. After training each of the classifier, cross validation was performed to validate and evaluate the performance of the classifier. The support vector machine gave the best classification result on my dataset of choice.

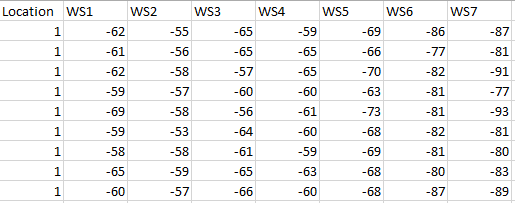
**2. Introduction**

Pattern recognition is the science of making inferences from perceptual data, using tools  
from statistics, probability, computational geometry, machine learning, signal processing,  
and algorithm design. A Pattern Recognition system is a system that is developed to  
perform pattern recognition automatically. In this project we develop a pattern  
recognition system to efficiently classify a real world dataset.

The project report has been structured in the following manner . At first I have presented  
the details of the datasetI am using including the pre-processing and feature extraction I  
have performed on he dataset. The next section gives the implementation details of each  
classifier with the results obtained. Lastly a comparative study has been performed on the  
classification results obtained by running each classifier on the given real world data set.

**3. Description of the dataset and Pre-processing**

The initial step in implementing the project was to choose a proper dataset to perform the  
classification . In order to do this I did my initial research on the datasets provided and  
chose the “Wireless Indoor Localization” dataset for my project. The dataset has around 200 data samples and 7 features. The data was collected from an indoor WiFi system which has 8 Wifi routers at various locations. column 2~8 refer to the measured signal strength (dB, integers) at 7 wireless sensors (WS1-WS7 below) (routers); column 1 is the user location (class). There are 500 data points for each location (class). Thus, there are 7 input variables (features) and 4 classes (user locations).



Also as part of preprocessing , I have tried to implemented the normalization process on the features but there is no better classification results. The normalization was done based  
on the standard scaler(such that its distribution will have a mean value 0 and standard deviation of 1).