Name(s)

Manogna Vemulapati

Project Number

35186

Project Title

Classification of ECG Arrhythmia Using Manifold Learning and Support Vector Machine

Abstract

Objectives/Goals

The purpose of my project is to design a computational method for efficient and accurate classification of ECG arrhythmia in a machine learning framework. The engineering goal of my project is to train a machine learning classifier on a dataset of ECG arrhythmia and then use the trained classifier to detect the presence or absence of arrhythmia in a given ECG test vector and classify it into one of 16 known categories.

Methods/Materials

Materials: Python programming environment, Scikit-Learn (a machine learning toolbox in Python), Python packages such as matplotlib, SciPy, NumPy and LinAlg, and arrhythmia dataset from UCI Machine Learning Repository.

Method:The main steps are: (a) Preprocess the data by standardizing (zero mean and unit variance for each feature), normalizing (unit norm for each feature vector) (b) Estimate the intrinsic dimension d of the manifold by plotting the manifold reconstruction error for each possible dimension of the manifold and calculating the elbow point at which the error flattens out (c) learn the lower d dimensional manifold representation of the data by using one of manifold learning algorithms such as Locally Linear Embedding (LLE) (d) train a SVM classifier on the lower dimensional dataset.

Results

Multi-class SVM classification of multiple types of arrhythmia has accuracy of 70%. By using the linear dimensionality reduction method Principal Component Analysis (PCA), the number of features have been reduced to 146 from 352 but the SVM binary classification accuracy is still as good as basic SVM at 90%. Manifold Learning method Locally Linear Embedding (standard, modified and ltsa LLE) performed the best by reducing the features to 128 and the subsequent SVM binary classification accuracy is 94%.

Conclusions/Discussion

Firstly, the best performance algorithm, LTSA, yielded a 62.5% reduction in the number of dimensions for the feature vectors, as well as a 3% increase in the accuracy score. This proves a significant improvement in accuracy and more importantly, saw a 20% increase in precision. When extrapolated to a real world scenario, the size of the dataset for each arrhythmia class would be much large and the use of manifold learning with the SVM classifier would likely be more successful. The increase in precision proves the existence of a nonlinear manifold for this dataset.

Summary Statement

My project uses a machine learning framework with dimensionality reduction to more accurately classify ECG arrhythmia.

Help Received

Project conducted with help of Computer Science teacher, Mr. Steinke.