matplotlib.pyplot

Predicting the Presence of Cardiac Arrhythmia

Stephanie Campa and Kathryn Bunner
{scampa, kbunner}@stanford.edu
Department of Computer Science, Stanford University

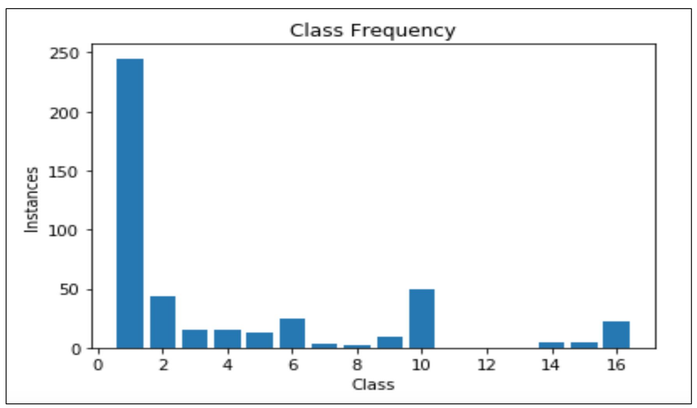
Abstract

Cardiac arrhythmias are caused by the malfunction of the electrical impulses that coordinate heartbeats, yet a patient can experience arrhythmia without noticing any symptoms. Using a variety of data including attributes like EKG results, we built a classifier that will predict either the presence or absence of cardiac arrhythmia. Because there are at least fourteen different classes of cardiac arrhythmia we tried a variety of experiments including logistic regression, support vector machines, and neural networks and were ultimately able to demonstrate improvements in classification against a baseline regression.

Data and Features

The dataset comes from the UCI Machine Learning Repository, and was donated by H. Altay Guvenir from Bilkent University. [1] It is a multivariate set that contains 452 instances and 279 different attribute with the majority being results from an EKG, but includes other attributes like age, sex, and weight.

Each instance is classified into one of 16 groups, where class 1 denotes no arrhythmia, classes 2-15 denote different types of arrhythmia, and unclassified instances are assigned to 16. The distribution is highlighted below.



Models

Preprocessing:

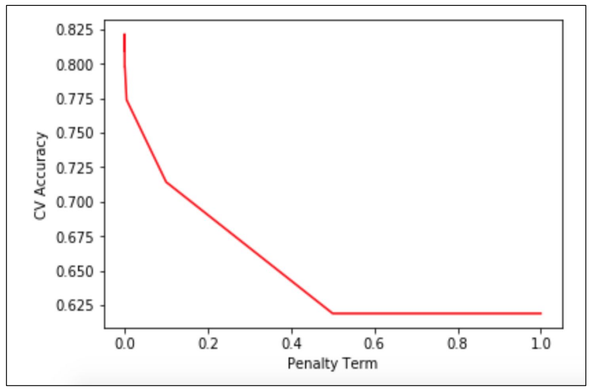
For our binary classification task of predicting the presence or absence of cardiac arrhythmia, we grouped all classes together where cardiac arrhythmia was present. For our multi-class tasks using the SVM and NN, we condensed our 16 classes down to 6, to ensure that each class contained enough data.

Logistic Regression:

For our baseline, we used the results of previous studies, as well as implemented a binary logistic classifier to detect whether any class of cardiac arrhythmia was present.

Support Vector Machines:

A Linear Kernel outperformed both the Radial Basis Kernel and the Sigmoid Kernel in both the binary and multi-class classification. The penalty term was set to 0.0001 and gamma was set to 1/n_features.



Accuracy vs Penalty Term of the Linear SVM for binary classification

Neural Networks:

We implemented a Multi-Layer Perceptron classifier to train on both our binary and multi-class data. The neural network utilized three hidden layers in addition in to the input and output layers, and used a learning rate of 0.01.

Results

	CV Accuracy	Test Accuracy
Logistic Regression	0.654	0.694
Binary Linear SVM	0.821	0.776
Binary RBF SVM	0.560	0.564
Multi Linear SVM	0.690	0.729
Multi RBF SVM	0.560	0.565
Binary Neural Net	0.693	0.741
Multi Neural Net	0.6.33	0.648

	Logistic Reg.	Binary SVM	Multi SVM	Binary NN	Multi NN
Precision	0.789	0.746	0.655	0.762	0.483
Recall	0.625	0.917	0.729	0.665	0.648
F1 Score	0.698	0.822	0.682	0.694	0.527

Discussion and Future Work

Using accuracy as our main measurement for evaluating our results, our models produced promising results for both arrhythmia prediction and arrhythmia classification. We were surprised the SVM outperformed the Neural Network because we thought the NN would be able to better detect the relevant features. Our model was overfitting due to the high number of features in our data, therefore, in the future more thoughtful data selection could produce higher accuracy and reduce the bias of our model.

References

[1] Guvenir, H. Altay, Acar, Burak, and Huderrisoglu, Haldun. (1997). UCI Machine Learning Repository [https://archive.ics.uci.edu/ml/datasets/ Arrhythmia]. Ankara, Turkey: Bilkent University, Department of Computer Engineering and Information Science.
[2] Guvenir, H. Altay; Acar, Burak; Demiroz, Gulsen and Cekin, Ayhan. (1997). "A Suupervised Machine Learning Algorithm for Arrhythmia Analysis." Bilkent University, Baskent University, Ankara, Turkey.
[3] Tang, Richard and Vyas, Saurabh. (2014) "Automated detection \& classification of arrhythmias." Stanford University, Department of Bioengineering.
[4] Savalia, Shalin, and Emamian, Vahid. (2018). "Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolutinoal Neural Networks." Bioengineering (Basel). 2018 Jun; 5(2): 35. 10.3390/bioengineering5020035
[5] https://scikit-learn.org/stable/modules/ generated/sklearn.svm.SVC.html#sklearn. svm.SVC.score
[6] Kingma, Diederik, and Jimmy Ba. “Adam: A method for stochasticoptimization.” arXiv preprint arXiv:1412.6980 (2014).