Evaluating Effectiveness of Differing Methods of Supervised Classification Algorithms on Non-Linearly Separable Datasets

Abstract

PURPOSE: Understanding the effectiveness of the Perceptron algorithm compared to the kernel PCA k means algorithm as a means for classifying complex data. Do these algorithms accurately and efficiently classify the data?

METHODS: Applying Perceptron Algorithm and Logistic Regression on a large non-linearly separable dataset, gathering data at distinct iteration numbers and comparing their accuracies. Then to test kernel PCA effectiveness, it is applied to the fisher iris data set, then compared to the original correct labels.

RESULTS: The perceptron algorithm proved to be unstable and inaccurate(87.6%) for non linearly separable data, scoring lower in accuracy compared to the logistic regression method (91.1%) accuracy. The kernel PCA k-means algorithm had a perfect 100% accuracy on the fisher iris data set.

CONCLUSIONS: The perceptron algorithm should not be used in order to correctly classify a non linearly separable dataset. Non linear algorithms like the logistic regression should be used in its place. The kernel PCA method is a very effective method for classification for this data and should be used for these non linearly separable datasets.

Introduction

Human brains are very good at classifying data. In an attempt to mimic human neurons, computers must be able to classify data efficiently and accurately. Classifying data can be done in many ways by computers, a staple of the method is the perceptron algorithm. The perceptron algorithm is a linear classification algorithm that uses iterations in order to brute force find an optimal weight vector that best classifies a hyperplane to separate two classifications. Each iteration changes the guessed hyperplane. The algorithm iterates until there is no more falsely classified data or until a maximum number of iterations occurs resulting in a linearly separated data set. However its behaviour is not defined for non linearly separated data, and that's what's getting explored now. We will be testing the effectiveness of the perceptron algorithm compared to a nonlinear classification method on a non linearly separable data set by measure of accuracy.

In addition, we will also inspect the effectiveness of the kernel PCA k-means method of classification of non linearly separable data. This method works by separating the data into a larger dimension where the data is linearly separable, then performing a k- means clustering to find the clusters. As transferring the data into a higher dimension is computationally expensive, it can be emulated by performing a “kernel trick” to simulate the data being in a higher dimension for the specific case of data separability. This data will be tested for effectiveness through its accuracy measure.

Methods

The college data was first loaded and prepared into data and labels. The data is then standardized and reduced down to 2 dimensions using PCA. The perceptron algorithm is then implemented, an initial weight vector is initialized. In each iteration, the error vector is calculated by scoring the data, gathering the correct labels, and subtracting the two, then updating the weight vector by adding a factor of the error vector. If the norm of the vector is nonzero and the iterations have not reached their max, then the iterations continue. Once this is done and a final hyperplane is obtained, it is then scored and the accuracy is obtained by putting the number of errors divided by the total observations. Then the logistic regression is obtained via a Matlab function, and its accuracy computed as well. The accuracies are then compared to each other. The ROC curves are calculated using the Matlab function and plotted as well.

In order to test the effectiveness of the kernel PCA k-means algorithm, the fisher iris dataset is loaded and prepared into data and labels. The Gram matrix is calculated via the kernel equation that is found within the course notes (33.14). The position (i,j) in the center "scatter matrix" of the gram matrix is gathered by applying the kernel function to the original points. The kernel function is described as the distance between observation I and observation J squared then multiplied by constants and put as the exponent of *e.* The matrix is iterated through in order to create every point then once created it is centered via the Gmat Matlab function.

Once the Gram matrix is created, its eigenvectors are extracted and sorted in a descending order. In order to do PCA, the largest 2 eigenvectors are extracted and scored in order to create a 2 dimensional version of the data. Kmeans is applied to the 2 dimensional dataset by the Matlab function. Its accuracy is calculated using the given labels and the k means clustering.

Results

Table 1: Displays the accuracies of each of the algorithms.

Method	Accuracy
Perceptron Algorithm (10000)	0.8764
Perceptron Algorithm (30000)	0.8867
Perceptron Algorithm (50000)	0.8533
Logistic Regression	0.9112
Kernel PCA	1

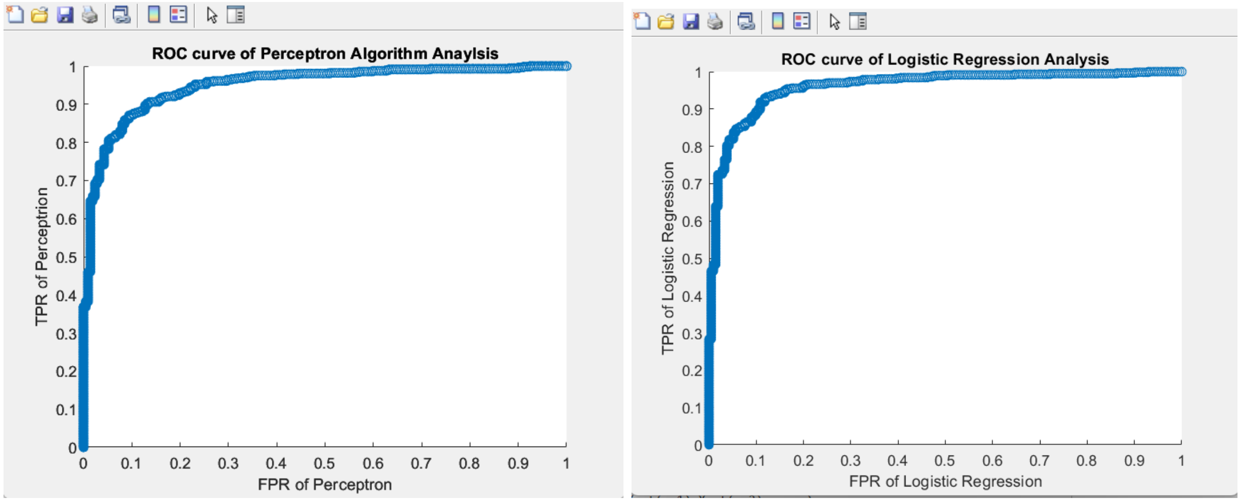


Figure 1 and 2: ROC curve of Perceptron Algorithm and Logistic Regression

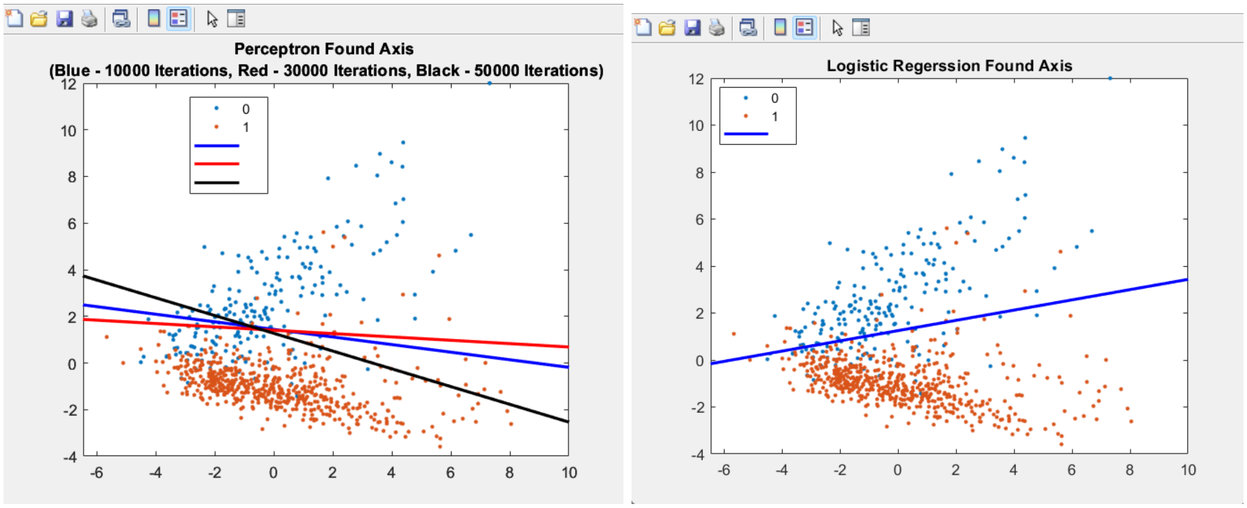
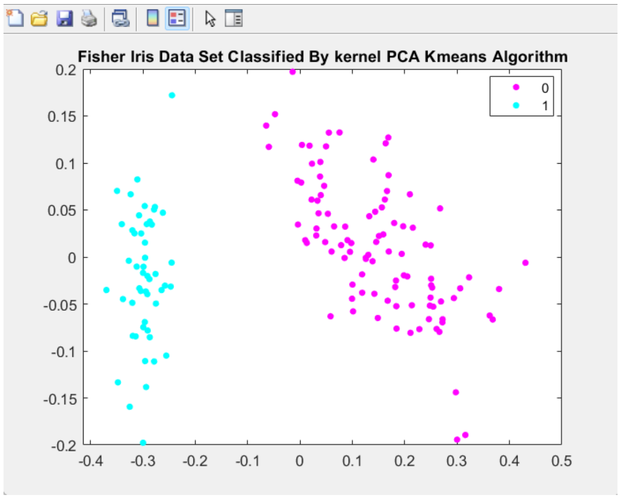


Figure 3 and 4: Visualization of the dimensionality reduced data separated by the corresponding hyperplanes,



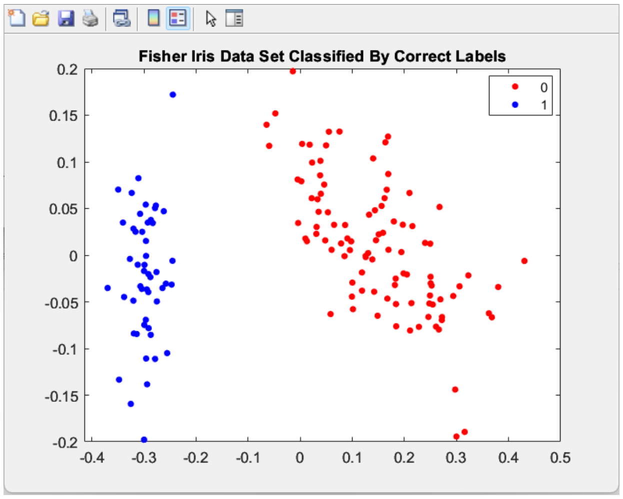


Figure 5 and 6: Scatter Plots depicting the results of the kernel PCA k means classification, and the true labels of the fisher iris dataset.

Discussion

Looking at the figures 3 and 4, the scatterplots with the corresponding hyper planes (Only the 10000 iterations value). Visually, the data separated by the perception did a worse job separating the data, but not a bad separation, evident in the decent accuracy of 87%, much better than random guessing. This is also reflected in the incredibly high ROC curve that corresponds to a large AUC value, an indicator of a good classification. This small inaccuracy can be due to the duty of the perceptron algorithm, its supposed to find a separability in linearly separable data, if the data does not converge to a correct point where there are no incorrect answers, the data will iterate forever causing a phenomenon called “wobbling”. This happens in which the hyperplane cannot converge to a point and will overcorrect, becoming less accurate. This is evident in the 2 other hyperplanes, the ones at 30000 and 50000 iterations. The model gets more accurate at 30000, at approximately 89% accuracy, but at 50000 iterations, the model becomes less accurate at 85%. This shows the major flaw in the Perceptron, the data is unreliable if it does not converge to a specific model, however if it does it is a very good model. This is less so the case with the logistic regression, as it's an algorithm designed for non linearly separable data, it performed much better separating the data, however the difference was not too large. There was only a 3.5% increase in performance compared to the 10000 iteration perceptron. The data confirms that the perceptron is not suited for non linearly separated data and a different algorithm should be used in its place if the data is found to be non linearly separable.

As the kernel PCA was done on a separate dataset, it cannot be compared to the Perceptron or the logistic regression. Visually inspecting the resulting grouped data, there are no differences between the correct labels and the kmeans clustered data. This means that the k-means algorithm is a perfect option for this data set. This leads to the conclusion that this data set can be linearly separated in higher dimensions and that the Kernel PCA K-means clustering is an efficient and effective classification algorithm for this data.