A Report for Foundations of Machine Learning Lab 2

Name：Ziyi Guo
Email：zg2u21@soton.ac.uk
School of Electronics and Computer Science

1 General Implement of Perceptron Algorithm

In this part, samples from two bi-variant Gaussian densities with distinct means m_1 = $\begin{bmatrix} 0 \\ 5 \end{bmatrix}$ and m_2 = $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$, and identical covariance matrices C = $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ were generated and used to draw the subplot of their distribution (Figure 1).

Then, a perceptron algorithm was trained to classify the data. Basically, a perceptron decision function assigns the data to one class or the other depending on whether $w^T x \lessgtr 0$, in which x and w represents the input data and the weights. For random index τ, if $w^T x^{(\tau)} \leqslant 0$, w will be iterated through

$$w^{(\tau+1)} = w^{(\tau)} + \alpha\, y^{(\tau)} x^{(\tau)}$$

It is the goal to driving the scalar product of $w^T x^{(\tau)}$ and target $y^{(\tau)}$ being positive over all the data in the end. To achieve this, data from the two classes was concatenated and randomly partitioned into training and test sets, labels {1, -1} were set to indicate the two classes, the iterative error correcting learning was carried out, and the times of iteration and the corresponding classification accuracies were worked out and used to plot the learning curves in the figure below:

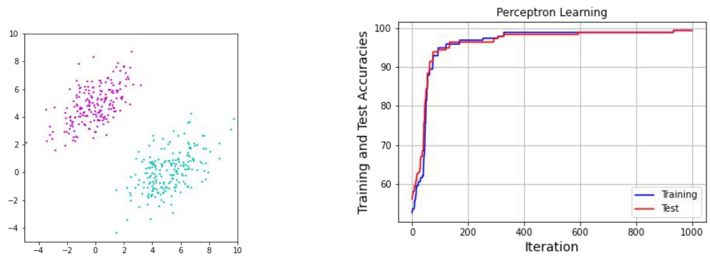

Figure 1. Subplot of Training Data from two Bivariant Gaussian Density and Learning Curves for its Classification

Here the perceptron tool in the scikitlearn package was also implemented to make a comparison on the algorithms' performances. The accuracies on training set and test set are 1.00 and 0.99.

2 A Specific Problem to Test Perceptron Algorithm Performance

In this part, problem of classifying samples from two Gaussian densities with means at m_1 = $\begin{bmatrix} 2.5 \\ 2.5 \end{bmatrix}$, m_2 = $\begin{bmatrix} 10.0 \\ 10.0 \end{bmatrix}$ and identical covariance matrices C = $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ was raised. The same perceptron was used to classify the samples. The corresponding learning curves are as below:

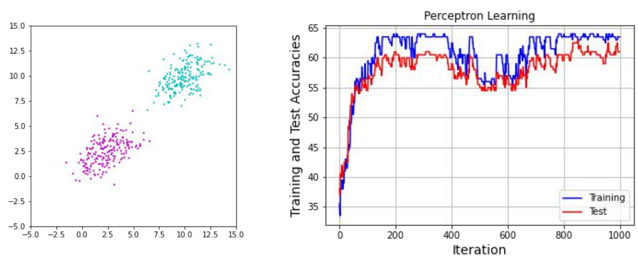

Figure 2. Subplot for New Samples and the New Perceptron Learning Curves

It is observed that the learning curves cannot converge and the accuracies are quite low. However, once the data was transformed into a higher dimension, through data appending operation, correspondingly altering parameter *w* to a three-column vector, the situation changed. The accuracies of the new algorithm increased into 100.00 and 98.00 at the end of the training (Figure 4). It is proved that we could use methods of increasing data dimensions to make the data separable in high dimension and improve the performance of perceptron.

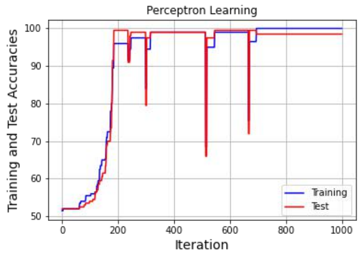

Figure 3. Learning Curves for High-Dimension Perceptron

3 Perceptron for Indoor User Movement Prediction from RSS Dataset

In this part, the chosen dataset represents a real-life benchmark in the area of Ambient Assisted Living applications. The binary classification task is predicting the pattern of user movements in real-world office environments from time-series generated by a Wireless Sensor Network (WSN).
Target data consists in a class label indicating whether the user's trajectory will lead to a change in the spatial context (i.e. a room change) or not. In particular, the target class +1 is associated to the location changing movements, while the target class -1 is associated to the location preserving movements. Each input file in the provided dataset contains data pertaining to one temporal sequence of input RSS data (1 user trajectory for each file). The dataset contains 314 sequences, for a total number of 13197 steps.
After necessary data cleansing, data sequences were transformed to feature matrix in 4 cols. The result of using Perceptron to classify the data is in the figure below. It is observed that the accuracy of Perceptron is quite low and the Learning Curve cannot converge, indicating that it is not the suitable method to deal with classification problem with complicated features.

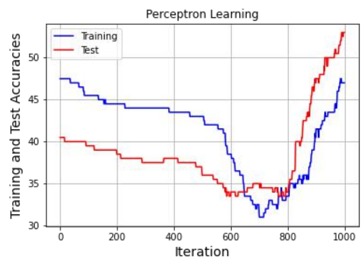

Figure 4. Learning Curves for Perceptron Classifier on User Movement

Comparatively, related essay on the dataset purposed the Echo State Network(ESN) based on RNNs and reached an accuracy of 0.92[1]. It is proved that building a neural network representing the users' behavior is crucial to improve the performance of the classifier.

Reference

[1] Gallicchio C, Micheli A, Barsocchi P, et al. User Movements Forecasting by Reservoir Computing Using Signal Streams Produced by Mote-Class Sensors[C]//International Conference on Mobile Lightweight Wireless Systems. 2012.