Cognitive Algorithms Assignment 2

Handwritten Digit Recognition

Due on Thursday, May 26, 2016 , 10 am via ISIS

In the lecture, you learned about the perceptron and the prototype classifier, which is also called the nearest centroid classifer (ncc). Both are linear classification methods. Given training data

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n) \in \mathbb{R}^d \times \{-1, 1\},$$

their goal is to learn a weight vector $\mathbf{w} \in \mathbb{R}^d$ and a bias term $b \in \mathbb{R}$, such that each new data point $\mathbf{x} \in \mathbb{R}^d$ will be assigned the correct class label via the following function:

$$\mathbf{x} \mapsto \text{sign}(\mathbf{w}^T \cdot \mathbf{x} - b).$$

The linear perceptron and the prototype classifier use different strategies to achieve this goal. In this assignment you will compare the two methods to predict handwritten digits. The task is to classify **one** digit against all others.
Download the python template `assignment2.py` and the data set `usps.mat` from the ISIS web site.

1. **(1 point)** The data set `usps.mat` contains handwritten digits from the U.S. Postal Service data set. Familiarize yourself with the data by loading the data and plotting some images, with the provided functions `load_usps_data` and `plot_imgs`.
   Answer the following questions:
   How many images does the data set contain? Of how many pixels does each image consists of?

2. **(10 points)** Implement a linear perceptron by completing the function stub `train_perceptron`. Call the function `analyse_accuracies_perceptron` for a digit of your choice. It plots the classification accuracy, i.e. the percentage of correctly classified data points, as a function of iterations as in Figure 1. Does the accuracy converge? Yes, it does converge.

3. **(8 points)** Implement a prototype/nearest centroid classifier by completing the function stub `train_ncc`. Note that points will be deducted for the use of loops.

4. **(6 points)** Complete the function stub `plot_histogram` that calculates the classification accuracy and plots a histogram of classifier output $\mathbf{w}^T\mathbf{x}$ for each class.

5. **(3 points)** Call `compare_classifiers` for a digit of your choice. It plots, for both the perceptron and the nearest centroid classifier, the histogram of classifier outputs and the weight vector as in Figure 2. Call the function several times: which algorithm (Nearest Centroid Classifier or Perceptron) would you prefer for this task? Perceptron is better

6. **(2 points)** In this task, you have trained and tested the algorithm on the same data set. Is this enough to reach a final conclusion on which algorithms performs good on this type of data? No

After completing the tasks, hand in your completed `assignment2.py` via ISIS. Please write your name and your Matrikel Number as the first line of the code. Also hand in a pdf file that contains your name, the answers to the questions and two plots (the plots similar to Figure 1 and 2).

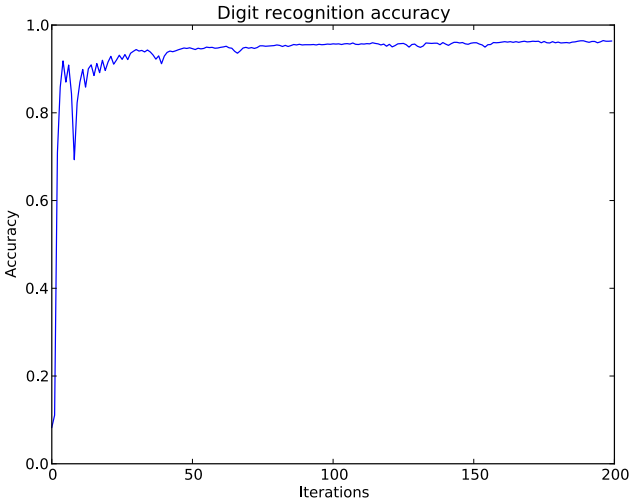

Figure 1: The perceptron on digit 3. Classification accuracy as function of iterations.

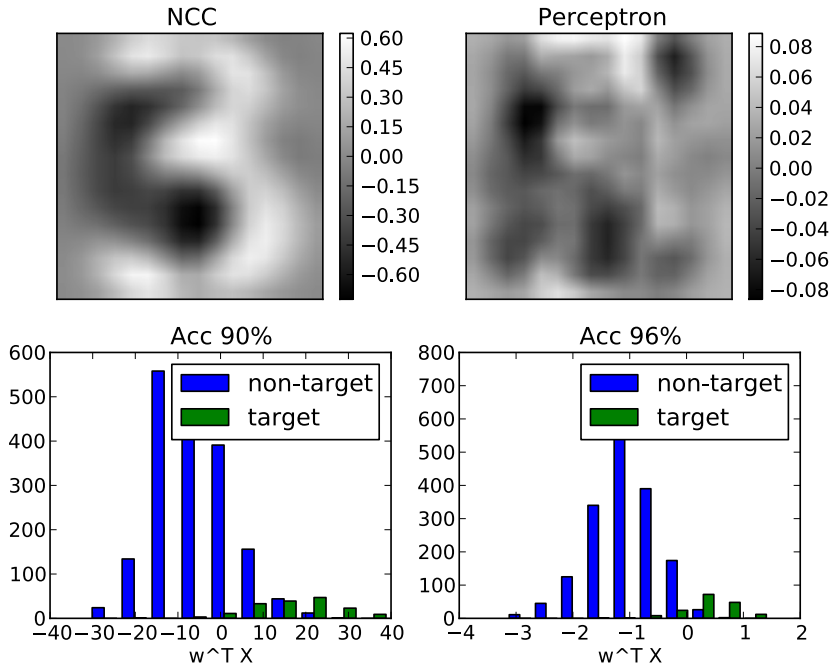

Figure 2: Classification results for digit 3. (Upper left) Weight vector of nearest-centroid-classifier. (Upper right) Weight vector of perceptron. (Lower left) Histogram of classifier output of nearest-centroid-classifier. (Lower right) Histogram of classifier output of perceptron.