

TDA 231 Machine Learning: Homework 2

Instructor: Chiranjib Bhattacharyya and Devdatt Dubhashi

Due Date: February 7, 2012

Goal: Fisher Discriminant, Perceptron Algorithm

General guidelines:

1. All datasets can be downloaded from the course website.
2. All matlab files have to be submitted as a single zip file named *code.zip*.
3. All plots, tables and additional information should be in a single pdf file named *report.pdf*.
4. The report should clearly indicate your group number on the fire system, your names, personal numbers and email addresses.

Useful matlab functions:

- *General:* randperm

Both datasets used in this assignment have two variables X and Y , where each row of X is a two-dimensional data point while the corresponding element in Y is the label (-1 or $+1$) for that data point.

1. (2 points) Implement Fisher's discriminant on *dataset3.mat*. Report the classification error (number of mis-classified points using the obtained classifier) and the classifier $[w, b]$.
2. (2 points) Implement Fisher's discriminant on *dataset4.mat*. Report classification error as well as $[w, b]$.
3. (3 points) Implement perceptron algorithm on both the datasets.
 - (a) For each of the datasets plot the number of updates in x -axis and $\|w\| = \sqrt{\sum_{i=1}^d w_i^2}$ in the y -axis. You will have to submit two graphs.
 - (b) Redo previous question (a) with a random (permuted) update order. Submit the plots for these.
 - (c) (Optional - Not to be submitted) When does the perceptron algorithm converge? How does this relate to $\|w\|$ as algorithm progresses?
4. (3 points) Re-implement Fisher's discriminant and Perceptron algorithm on *dataset4.mat* with the following feature map $\phi(x) = [x_1, x_2, x_1^2, x_2^2]$. You will need to design your classifier based on $\phi(x)$.
 - (a) Report classification error for both classifiers.
 - (b) Submit the plot of $\|w\|$ versus number of updates.
5. (Optional - Not to be submitted) Using one of the feature maps discussed in assignment one, compare Fisher's discriminant with perceptron algorithm on the *digits.mat* dataset.