ECSE-6610: PR Homework Set 3

Chengjiang Long

March 6, 2018

Assigned Date: Feb 27, 2018.
Due Date: Mar 9, 2018.
Collaboration Policy. Homeworks will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited.
Late Policy. No late submissions will be allowed without consent from the instructor. If urgent or unusual circumstances prohibit you from submitting a homework assignment in time, please e-mail me explaining the situation.
Submission Format. Electronic submission of a zip file is mandatory. Include code in your pdf file as needed to make your answers clear. Submit all code separately.

Problem 1 (20 points) Consider the hyperplane used for discriminant functions.

(a) [**10 points**] Show that the distance from the hyperplane $g(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_0 = 0$ to the point $\mathbf{x}_a$ is $|g(\mathbf{x}_a)|/\|\mathbf{w}\|$ by minimizing $\|\mathbf{x} - \mathbf{x}_a\|^2$ subject to the constraint $g(\mathbf{x}) = 0$.

(b) [**10 points**] Show that the projection of $\mathbf{x}_a$ onto the hyperplane is given by

$$\mathbf{x}_p = \mathbf{x}_a - \frac{g(\mathbf{x}_a)}{\|\mathbf{w}\|^2}\mathbf{w}$$

Problem 2 (20 points) Consider the following 5D dataset in which the first column is the class label:

ω_2: (1, 1, -1, 0, 2)
ω_1: (0, 0, 1, 2, 0)
ω_2: (-1, -1, 1, 1, 0)
ω_1: (4, 0, 1, 2, 1)
ω_1: (-1, 1, 1, 1, 0)
ω_1: (-1, -1, -1, 1, 0)
ω_2: (-1, 1, 1, 2, 1)

(a) [**5 points**] State the desired condition for correct classification for each sample using a linear discriminant function, before and after normalization.

(b) [**15 points**] Train a perceptron using the single sample rule with the learning rate kept at 1.0 for all iterations. Use $[3, 1, 1, -1, 2, -7]$ as the initial weight vector. Make sure that the first element of the weight vector corresponds to class label. Show all steps.

Problem 3 (20 points) Consider the sum-of-squared-error criterion function

$$J_s(\mathbf{a}) = \sum_{i=1}^{n} (\mathbf{a}^t \mathbf{y}_i - b_i)^2$$

Let $b_i = b$ and consider the following six training points:
ω_1 : (1, 5), (2, 9), (-5, -3)
ω_2 : (2, -3), (-1, -4), (0, 2)
(a) [**10 points**] Calculate the Hessian matrix for this problem.
(b) [**10 points**] Assuming the quadratic criterion function calculate the optimal learning rate η.
(*Hint: you can indicate the optimal learn rate η as an expression in term of both* **a** *and b if necessary. If you feel difficult in calculating gradients and Hessian matrix, I would like to suggest you to read the online Duda's textbook, Page 286, Equation 42-43.*)

Problem 4 (40 points + Extra 20 points) Download the "Pima Indians Diabetes Database" from http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes. Use a 50%-50% random split of the data for training and testing.

(a) [**5 points**] Apply Principal Component Analysis to reduce the dimensionality of the data from 8 (do not forget to exclude the class label before doing PCA) to 3. Explain how you selected the appropriate principal components.

(b) [**5 points**] Train a classifier using MLE after the data have been projected. Report average classification accuracy over at least 10 runs and the three principal components you selected for one of the runs.

(*Hint: the purpose of running multiple runs is to reduce the impact of random data split, and the average classification accuracy can reflect a more general performance.*)

(c) [**5 points**] Now apply the Fisher Linear Discriminant method to the Pima Indians Diabetes database. Use all 8 features, excluding the class label. Train a classifier using MLE after the data have been projected. Report average classification accuracy over at least 10 runs and the optimal projection direction for one of the runs.

(d) [**25 points**] Train a linear SVM classifier with 5-fold cross-validation using the same feature vectors in (b) and also train another linear classifier using the feature vectors in (c). Report average classification accuracy over at least 10 runs and the optimal projection direction for one of the runs.

(e) [**Extra 20 points for open solutions**] Now apply the kernel SVM classifiers with multiple kernels (should include both linear and nonlinear kernels) and compare their performances with the 1-NN classifier, as well as the performance in (b) to (d). Report average classification accuracy over at least 10 runs and the optimal projection direction for one of the runs.