HW-2

Monday, October 24, 2022 12:29 AM

HW-2

CS 559: Machine Learning

Homework Assignment 2

Due Date: Thursday 6:30 PM, October 20, 2022 **Total: 100 points**

Problem 1. Linear Discriminant Analysis (20 points)

Please download the Iris data set from the UCI Machine Learning repository and implement Linear Discriminant Analysis for each pair of the classes and report your results. Note that there are three (3) class labels in this data set. To implement these models, you can use python and the sklearn packages. Please submit the code along with each step of your solutions to get full points.

Link to the data: https://archive.ics.uci.edu/ml/datasets/Iris

Problem 2. Gradient Descent Algorithm and Logistic Regression (40 points)

(1) In logistic regression method, please derive the derivative of the negative logarithm of the likelihood function with respect to parameter w. You need to show the detailed steps to obtain the following results.

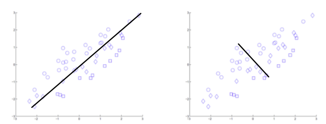
$$\nabla_w \varepsilon(w) = \sum_{n=1}^{N}(f(x_n) - y_n)x_n$$

(2) Please download the breast cancer data set from UCI Machine Learning repository. Implement your Logistic regression classifier with ML estimator using Stochastic gradient descent and Mini-Batch gradient descent algorithms. **Do not use any package/tool.** Use cross-validation for evaluation and report the recall, precision, and accuracy on malignant class prediction (class label malignant is positive). Write down each step of your solution.

Link to the data: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29

Problem 3: Principal Component Analysis (PCA) (20 points)

(1) Given labels of the data, the goal of Fisher's Linear Discriminant is to find the projection direction that maximizes the ratio of between-class variance and the within-class variance. While PCA aims to reduce the dimension of the data by finding projection directions that maximizes the variance after projection. Note that PCA does not consider the label information. In the following figures, consider round points as positive class, and both diamond and square points as negative class. Please draw (a) the direction of the first principal component in the left figure by ignoring the label of the data points, and (b) the Fisher's linear discriminant direction in the right figure. Please draw a line to show the direction for each of them.



(2) Consider 3 data points in the 2D space: (2,2), (0,0), (-2,-2). Please answer the following questions.

a) Calculate the first principal component by calculating the eigenvalue (non-zero) and eigenvector of the covariance matrix. You need to provide the actual vector of the first principal component (with length=1). You can use the unbiased estimation of the covariance:

$$var(X) = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})^2$$

$$Cov(X,Y) = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})(Y_n - \bar{Y})$$

b) If we project the three data points into the 1D subspace by the principal component obtained in (a), what are the new coordinates of the three data points in the 1D subspace? What is the variance of the data after projection?

c) What is the cumulative explained variance of the first principal component? Is there any variance that is not captured by it?

mean $=0$ $\frac{8}{2} = 4$ var X 4 is also covariance

4 var Y

$\begin{bmatrix} \text{Var} & \text{Covarian} \\ \text{Covar} & \text{Var} \end{bmatrix}$ $(4-\lambda)^2 - 16 = 0$

a. $\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$ $\lambda^2 - 8\lambda = 0$

$\lambda(\lambda - 8) = 0$

$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\lambda = 8$ $-4x_1 + 4x_2 = 0$ $x_1 = x_2$

$4x_1 - 4x_2 = 0$ $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

b. origin point minus mean (0)

$(2,2)(1,1) = 4$ mean $= 0$ $\frac{16+16}{2}$

$(0,0)(1,1) = 0$ variance $= 16$

$(-2,-2)(1,1) = -4$

C. explained variance $= \frac{8}{8+0} = 1$

so all variance is captured by it.

Problem 4: Support Vector Machines (20 points)

Given 10 points in Table 1, along with their classes and their Lagrangian multipliers (α_i), answer the following questions:

(1) What is the equation of the SVM hyperplane $h(x)$? Draw the hyperplane with the 10 points.
(2) What is the distance of x_6 from the hyperplane? Is it within the margin of the classifier?
(3) Classify the point $z = (3,3)^T$ using $h(x)$ from above.

Table 1: Data set for question 4

Data	x_{i1}	x_{i2}	y	α_i
x_1	4	2.9	1	0.414
x_2	4	4	1	0
x_3	1	2.5	-1	0
x_4	2.5	1	-1	0.018
x_5	4.9	4.5	1	0
x_6	1.9	1.9	-1	0
x_7	3.5	4	1	0.018
x_8	0.5	1.5	-1	0
x_9	2	2.1	-1	0.414
x_{10}	4.5	2.5	1	0

1. $W = (W_1, W_2)$

$W_1 = 1 \cdot 0.414 \cdot 4 + (-1 \cdot 0.018 \cdot 2.5) + 1 \cdot 0.018 \cdot 3.5 + (-1 \cdot 0.414 \cdot 2)$

$= 0.846$

$W_2 = 1 \cdot 0.414 \cdot 2.9 + (-1 \cdot 0.018 \cdot 1) + 1 \cdot 0.018 \cdot 4 + (-1 \cdot 0.414 \cdot 2.1)$

$= 0.3852$

$b_1 = (0.846)\ (0.3852)\ (2.1) = -3.50108$

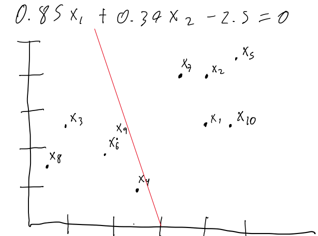
$b_2 = (0.846)\ (0.3852)\ (2.1) = -1.50092$

$b_1 = (0.846)\ (0.3852)\ (4) = -3.50108$

$b_2 = (0.846)\ (0.3852)\ (1) = -1.5002$

average $= -2.5$

$0.85x_1 + 0.39x_2 - 2.5 = 0$



2. $x_6 = (1.9, 1.9)$

$d = \frac{0.85(1.9) + 0.39(1.9) + (-2.5)}{\sqrt{(0.85)^2 + (0.39)^2}} = 0.15398$

$\frac{1}{\sqrt{0.85^2 + 0.39^2}} = 1.069$ margin so in it is in margin

3. $W \cdot X + b = 0.85$ It falls a little after the hyperplane

$3 + 0.39$ so X belongs to class label 1

$3 - 2.5 = 1.22$