MATH 3333 3.0 - Winter 2019-20
Assignment 3
(Due Date: March 12, 2020)

Question 1: Given the following data, please apply the Fisher linear discriminant method. There are two classes, C_1 and C_2. The class C_1 has five observations:

$$\begin{pmatrix} 2 & 3 \\ 3 & 5 \\ 4 & 4 \\ 5 & 6 \\ 6 & 6 \end{pmatrix}.$$

The class C_2 has six observations:

$$\begin{pmatrix} 2 & 0 \\ 3 & 2 \\ 4 & 2 \\ 4 & 3 \\ 6 & 4 \\ 7 & 6 \end{pmatrix}.$$

a) Comute the mean of the first class μ_1, and the mean of the second class μ_2.
b) Compute the within class variation $S_w = S_1 + S_2$, where S_1 and S_2 are the variations within C_1 and C_2, respectively.
d) Find the optimum projection v which can lead to the maximum seperation of the projected observations.
e)Find the cutoff point $\frac{1}{2}v^T\mu_1 + \frac{1}{2}v^T\mu_2$.
g) Given a new observation $(5.5, 5.5)$, which class does it belong to?

Question 2: In the forensic glass example, we classify the type of the glass shard into six categories based on three predictors. The categories are: WinF, WinNF, Veh, Con Tabl and Head. The three predictors are the mineral concentrations of Na, Mg, and Al. Attached is the R output of the multinomial logistic regression. The R function vglm considers the last group as the baseline category. The estimates of the five intercepts and the estimates of the 15 slopes are provided in the output. The model contains 20 parameters, which are estimated on 214 cases.
a) Let p_{ij} denote the probability that the ith observation belongs to class j. Formulate the logistic model for the five log odds: $\log \frac{p_{i1}}{p_{i6}}$, $\log \frac{p_{i2}}{p_{i6}}$, $\log \frac{p_{i3}}{p_{i6}}$, $\log \frac{p_{i4}}{p_{i6}}$, $\log \frac{p_{i5}}{p_{i6}}$.

b) The ith piece of glass shard is obtained and the Na, Mg, Al concentrations are: 0.15, 0.03, and 0.11, respectively. Calculate the probabilities p_{i1}, p_{i2},p_{i3}, p_{i4}, p_{i5}, and p_{i6}. Based on the predicted class probability, which type of glass does this piece of glass belong to?

Question 3:

```
Coefficients:
               Estimate Std. Error   z value
(Intercept):1  1.613703    0.84001  1.921057
(Intercept):2  3.444128    0.72131  4.774792
(Intercept):3  0.999448    0.93007  1.074594
(Intercept):4  0.067163    0.95554  0.070288
(Intercept):5  0.339579    0.89779  0.378239
Na:1          -2.483557    0.65323 -3.801955
Na:2          -2.031676    0.55399 -3.667326
Na:3          -1.409505    0.72721 -1.938243
Na:4          -2.382624    0.59434 -4.008837
Na:5           0.151459    0.53353  0.283879
Mg:1           3.842907    0.76674  5.012003
Mg:2           1.697162    0.47748  3.554387
Mg:3           3.291350    1.02370  3.215158
Mg:4           0.051466    0.50284  0.102351
Mg:5           0.699274    0.50346  1.388924
Al:1          -3.719793    0.68049 -5.466312
Al:2          -1.704689    0.54805 -3.110489
Al:3          -3.006102    0.75556 -3.978654
Al:4           0.263510    0.40013  0.658562
Al:5          -1.394559    0.51315 -2.717660

Number of linear predictors:  5

Names of linear predictors:

log(mu[,1]/mu[,6]), log(mu[,2]/mu[,6]), log(mu[,3]/mu[,6]), log(mu[,4]/mu[,6]),
log(mu[,5]/mu[,6])

Dispersion Parameter for multinomial family:   1

Residual deviance: 379.6956 on 1050 degrees of freedom

Log-likelihood: -189.8478 on 1050 degrees of freedom

Number of iterations: 7
```

a. In this question, we consider the discrimant analysis method for multivariate normal data. Given C_1, C_2, ..., C_K classes, we assign the prior probabilities to each class $P(C_j)$, $j = 1, \ldots, K$. Given that X belongs to class C_j, the conditional distribution of X is a multivariate normal with the mean μ_j, and the covariance matrix Σ_j. Then based on the Bayes formula,

$$p(C_j|X) = \frac{p(C_j)P(X|C_j)}{\sum_{j'=1}^{K} p(C_{j'})P(X|C_{j'})}.$$

Then we can use $P(C_j|X)$ as the discriminant function. We assign X to class j if $P(C_j|X) > P(C_{j'}|X)$, for any other classes. As the denominator is a constant which does not depend on j, we can use $P(C_j)P(X|C_j)$ as the discriminant function. Or equivalently we can use $\log P(X|C_j) + \log P(C_j)$. The discriminant function is denoted by $g_j(X)$.

$$\begin{aligned} g_j(X) =& logP(X|C_j) + \log P(C_j) \\ =& -\frac{1}{2}(X-\mu_j)^T\Sigma_j^{-1}(X-\mu_j) - \frac{1}{2}\log|\Sigma_j| + \log p(C_j) \end{aligned} \tag{1}$$

Consider the case that $\Sigma_j = \sigma^2 I$. In this case, all the predictors are independent with different means and equal variances σ^2. Please simplify $g_j(X)$ and show that it is a

linear function of X.

- b) In this example, we have three classes, each is a 2-dim Gaussian distribution, with $\mu_1 = (2,1)^T$, $\mu_2 = (4,4)^T$, $\mu_3 = (-2,3)^T$ and $\Sigma_1 = \Sigma_2 = \Sigma_3 = 2I_2$, where I_2 is an identity matrix of dimension 2×2. We assume the priors $p(C_1) = p(C_2) = 1/4$, and $p(C_3) = 1/2$. Let $X = (0.25, 0.5)^T$. Calculate $g_1(X), g_2(X)$ and $g_3(X)$. Classify the observation X to one of the classes.

Question 4: Analyze the German data set from the site: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data). Apply the linear discriminant analysis on the dataset. Please randomly select 800 observations as the training set and use your model to predict the default status of the remaining 200 loans. Repeat this cross-validation five times and calculate the average misclassification error.

Question 5:

Suppose we have 2-classes observations with p-dimensional predictors. We have samples $x_1, \ldots, x_n$, with n_1 samples from Class 1 and n_2 samples from Class 2. Let v be a unit vector. The projection of sample x_i onto a line in direction v is given by the inner product of $y_i = v^T x_i$. Let μ_1 and μ_2 be the means of class 1 and class 2. Let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the mean of the projections of class 1 and class 2. Denote the variance of the projected samples of class 1 is $\tilde{S}_1^2 = \sum_{x_i \in C1}(y_i - \tilde{\mu}_1)^2$ and the variance of the projected samples of class 2 is $\tilde{S}_2^2 = \sum_{x_i \in C2}(y_i - \tilde{\mu}_1)^2$. The Fisher linear discriminant is to project to a direction v which maximizes:

$$J(v) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Let the variance of the original samples of class 1 be $S_1^2 = \sum_{x_i \in C1}(x_i - \mu_1)((x_i - \mu_1)^T$. and the variance of the original samples of class 2 be $S_2^2 = \sum_{x_i \in C2}(x_i - \mu_2)((x_i - \mu_2)^T$. Define the within class variation:

$$S_w = S_1 + S_2.$$

Define the between the class variation: $S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$. Prove the objective function can be simplified as:

$$J(v) = \frac{v^T S_b v}{v^T S_w v}.$$