Quiz 4. Classification II: Logistic Regression (cont.), LDA, QDA
Name: Lisa Liubovich Attempt (circle one): BEFORE AFTER

1. If an unauthorized user enters the correct username and password, can this intrusion be detected? Modern methods of intrusion detection measure the time a user takes to enter a password, along with other factors, and compares them with those of the account owner. We will consider the time to enter the password in this exercise.
   Suppose:
   - An account owner spends 1.4 seconds, on average, to type their password.
   - Unauthorized users spend 2.7 seconds, on average, to type the same password.
   - In both groups, the times of typing passwords follow Normal distributions with the same standard deviation of 0.25 seconds.

   Now, it took someone 2.1 seconds to type the password.

   (a) Suppose the prior probabilities of authorized and unauthorized access are 0.5 and 0.5. Use discriminant analysis to conclude whether the account was accessed by an unauthorized user.
   Mu1 = 1.4
   Mu2 = 2.7
   Sd = 0.25
   Prior probabilities : P(authorized) = 0.5 and P(unauthorized) = 0.5
   Observed time: x = 2.1

   P(2.1|authorized) = 1/square root of 2pi * (0.25)^2 exp (-(2.1-1.4)^2/2(0.25)^2) = 1/square root of 2pi * 0.0625) exp (-0.7^2/2(0.25)^2) = 1/square root of 0.3927 exp (-0.49/0.125) = 1/0.6267 * exp(-3.92) = 1.5958 * 0.0198 = approximately 0.0316

   P(2.1|unauthorized) = 1/square root of 2pi * (0.25)^2 exp (-(2.1-2.7)^2/2(0.25)^2) = 1/= 1/square root of 2pi * 0.0625) exp (-0.6^2/2(0.25)^2) = 1/ square root of 0.3927 exp(-0.36/0.125) = 1/0.627exp(-2.88) = 1.5958 * 0.0561 = approx. 0.0895

   P(2.1) = P(2.1 | authorized) * P(authorized) + P(2.1 | Unauthorized) * P(Unauthorized) = = 0.0316 * 0.5 + 0.0895 * 0.5 = 0.0158 + 0.04475 = 0.06055

   P(authorized | 2.1) = (P(2.1| authorized) * P(authorized))/P(2.1) = (0.0316 * 0.5)/0.6055 = 0.0158/0.06055 = approx.. 0.261

   P(unauthorized | 2.1) = (P(2.1| unauthorized) * P(unauthorized))/P(2.1) = 0.0895 * 0.5/0.06055 = 0.04475/0.06055 = approx. 0.739

   Since the posterior probability of unauthorized access (0.739) is higher than that of authorized access (0.261), we can conclude that the account was likely accessed by an unauthorized user

(b) Does your classification correspond to LDA, QDA, or neither? Explain.

My classification corresponds to Linear Discriminant Analysis (LDA) because both the authorized and unauthorized users' typing times follow normal distributions with the same standard deviation of 0.25. LDA assumes that each class (in this case, authorized and unauthorized) has the same covariance matrix. In LDA, the decision boundary between classes is linear, which is appropriate here because the likelihood functions and the prior probabilities combine in a way that creates a linear decision boundary when the variances are equal.

(c) (**For Stat-627 only**) Redo question (a) with the consideration that: the prior probability of unauthorized use is 0.2, and the prior probability of authorized use is 0.8. (You can use R as a calculator. R function `dnorm(x, mean, sd)` computes the pdf (i.e., f(x)) of Normal distribution.)

P(unauthorized) = 0.2, P(authorized = 0.8)
P(x = 2.1 | authorized) = approx. 0.0316, P(x = 2.1| unauthorized) = 0.0895

P(authorized| x = 2.1) = 0.5857187 = approx. 0.586
P(unauthorized| x = 2.1) = 0.4142813 = approx. 0.414

Given these posterior probabilities, with the new pripor probabilities, we would conclude that the account was likely accessed by an authorized user.

(Stat-427, continue to next page for Q.2!)

2. (**For Stat-427 only**) Risk and loss function. In a binary (0 or 1) classification analysis, a model estimated that $\hat{\pi}_{Y=1} = 0.731$. Suppose the loss of a False Negative (a Y = 1 classified as $\hat{Y} = 0$) is three (3) times the loss of a False Positive (a Y = 0 is classified as $\hat{Y} = 1$). In view of this, will you classify the observation as 0 or 1?