Departamento de Matemática

Multivariate Analysis

Master in Eng. and Data Science & Master in Mathematics and Applications

2^{nd} Test - Part II — 1^{st} Semester – 2020/2021

Duration: 45 minutes — 04/02/2021 – 17:45

Please justify conveniently your answers

If the second letter of your first name is between "A" and "L" i solve **Group II - Version A**, otherwise solve **Group II - Version B**.

Any wrong choice of Group II Version will not be classified.

Group II - Version A **10.0 points**

1. An observation x comes from one of two populations with prior probabilities $P(Y = 0) = 1 - p$, $P(Y = 1) = p$, $0 < p < 1$, and probability density functions:

$$f_{X|Y=j}(x) = \begin{cases} \lambda\gamma_j(\gamma_j x)^{\lambda-1}e^{-(\gamma_j x)^\lambda}, & x > 0 \dots \\ 0, & \text{otherwise,} \end{cases}$$

with $\lambda > 0$, $\gamma_1 > \gamma_0 > 0$, $j = 0, 1$.

   (a) Derive, with statistical rigour, the classification rule that minimizes the total probability of misclassification. (4.0)

   (b) Assuming $\lambda = 1$, $\gamma_0 = 1$, $\gamma_1 = 10$, and $p = 0.4$ obtain and interpret:

      i. The sensitivity. (2.0)

      ii. The positive predictive value. (2.0)

Reminder: The distribution function of $X|Y = j$ is $F_{X|Y=j}(x) = 1 - e^{-(\gamma_j x)^\lambda}$, $x > 0$, $j = 0, 1$.

2. Comment the following sentence: (2.0)

"*The inclusion of irrelevant variables in Linear Discriminant Analysis leads to overfitting.*"

If the second letter of your first name is between "A" and "L" i solve **Group II - Version A**, otherwise solve **Group II - Version B**.

Any wrong choice of Group II Version will not be classified.

Group II - Version B **10.0 points**

1. An observation x comes from one of two populations with prior probabilities $P(Y = 0) = 1 - p$, $P(Y = 1) = p$, $0 < p < 1$, and probability density functions:

$$f_{X|Y=j}(x) = \begin{cases} \lambda\gamma_j(\gamma_j x)^{\lambda-1}e^{-(\gamma_j x)^\lambda}, & x > 0 \dots \\ 0, & \text{otherwise,} \end{cases}$$

with $\lambda > 0$, $\gamma_1 > \gamma_0 > 0$, $j = 0, 1$.

   (a) Derive, with statistical rigour, the classification rule that minimizes the total probability of misclassification. (4.0)

   (b) Assuming $\lambda = 1$, $\gamma_0 = 1$, $\gamma_1 = 10$, and $p = 0.4$ obtain and interpret:

      i. The specificity. (2.0)

      ii. The negative predictive value. (2.0)

Reminder: The distribution function of $X|Y = j$ is $F_{X|Y=j}(x) = 1 - e^{-(\gamma_j x)^\lambda}$, $x > 0$, $j = 0, 1$.

2. Comment the following sentence: (2.0)

"*Linear Discriminant Analysis is a generalization of Fisher's Discriminant Analysis.*"