



## TERCER PARCIAL

### Indicaciones generales

- Este es un examen se debe entregar el viernes 26 de mayo antes de las 10am.
- Las respuestas deben estar totalmente justificadas.
- ¡Suerte y ánimo!

Cargue la base de datos que está en e-aulas que tiene datos sobre diferentes vinos y su calidad.

1. (30 pts) La variable quality da la calidad de los vinos. Se dice que un buen vino es aquel que tiene más de 6 de nota en esa variable. El resto de vinos se consideran malos. Haga una clasificación utilizando discriminante de Fisher en el que clasifique entre buen vino y mal vino a partir de las demás variables.
  - a) ¿Cuál es la probabilidad previa estimada de pertenecer a cada una de las categorías?
  - b) Calcule la matriz de confusión y el APER.
  - c) ¿Cuál es la categoría más difícil de clasificar?
2. (15 pts) Divida la base de datos original en dos grupos de variables (los que quiera) para hacer dos particiones.
  - a) ¿Cuál es la correlación más alta entre los dos grupos elegidos?
  - b) ¿Qué variable influye más en cada grupo en esa correlación?
3. (2 pts) Defina con sus palabras el p-valor
4. (1,5 pts) ¿Cuál o cuáles de las siguientes afirmaciones es correcta?
  - a) Si hago un PCA puedo ver la correlación entre las variables, pero no me da información sobre la relación de las observaciones.
  - b) Después de preguntar durante 2 años la definición de p-valor, por fin todos tendrán bien esa pregunta porque tienen una semana para responder.
  - c) Millitos será campeón este año en Colombia.
  - d) Bad Bunny no sabe cantar.
5. (1,5 pts) Si hiciera una regresión lineal donde la variable respuesta es la cantidad de dinero que se gasta una persona en entradas a conciertos de Bad Bunny, y la variable explicativa es el gusto musical (numérica, siendo 0 muy mal gusto musical y 100 el mejor gusto posible). ¿El valor de  $\beta_1$  debería ser positivo o negativo? justifique su respuesta.