

Inferencia Estadística y Reconocimiento de Patrones - Segundo Cuatrimestre 2025

Trabajo Práctico 1 a ser resuelto con PYTHON y R

CONDICIONES FORMALES

- La consigna debe ser resuelta usando PYTHON y R, según corresponda.
- Fecha de entrega: hasta el miércoles 01/10 23 hs.
- Modalidad de trabajo: en pareja.
- Modalidad de entrega: cada pareja deberá subir **3** archivos a la tarea correspondiente en el aula virtual de la materia.

Archivo 1 en pdf: informe sobre la resolución del ejercicio. El informe detallará lo realizado en el ejercicio, demostrando la comprensión de los temas abordados.

Archivo 2 notebook implementada en lenguaje PYTHON, de extensión ipynb. Contendrá el código implementado para resolver parte del problema planteado y su nombre será **pareja0.ipynb** según el número de pareja que sea.

Archivo 3 implementado en lenguaje R. Contendrá el código implementado para resolver parte del problema planteado cuyo nombre será **pareja0.R** según el número de pareja que sea.

Es importante que tengan en cuenta que los archivos 2 y 3 no resuelven lo mismo. En el archivo 2 deben resolver lo que se pueda con PYTHON, y en el archivo 3, lo que se pueda con R. **Los archivos 2 y 3 son complementarios, no equivalentes.** Aunando los archivos 2 y 3 se debe encontrar una completa resolución del problema planteado.

El archivo **winequality-red.csv** contiene información sobre características de vinos tintos de Portugal. Se registraron las variables:

- fixed.acidity
- volatile.acidity
- citric.acid
- residual.sugar
- chlorides
- free.sulfur.dioxide
- total.sulfur.dioxide
- density
- pH
- sulphates
- alcohol
- quality

Para que el archivo sea correctamente leído en R, utilizar la siguiente línea de código:

```
vinos <- read.csv2("winequality-red.csv",dec=".")
```

Considerar a la variable respuesta **quality** y aplicar los siguientes métodos

■ de regresión:

- lineal múltiple.
- Ridge.
- LASSO.

■ de clasificación:

- regresión logística multinomial.
- K vecinos más cercanos.
- Bayes ingenuo.
- LDA.
- QDA.

1. Cuando se modele según la regresión lineal múltiple, decidir acerca de si la regresión es estadísticamente significativa e informar la significatividad de las variables explicativas.
2. En los casos que corresponda,
 - explicar lo realizado para la selección de parámetros mediante validación cruzada.
 - brindar los parámetros estimados.
3. Incluir gráficos que resulten apropiados.
4. En todos los casos evaluar métricas de los distintos métodos, para luego seleccionar el método, de regresión por un lado y de clasificación por otro, que mejor prediga la variable respuesta **quality**.
5. Elegir uno de los métodos de clasificación y explicar en profundidad el ajuste.
6. ¿Cuál de los ajustes realizados le parece el más adecuado para el conjunto de datos? Justificar.