

# Instrucciones Miniproyecto 2

## Aprendizaje Supervisado en Python

En este proyecto usted implementará modelos de aprendizaje supervisado. Dentro de la actividad, se reforzarán conceptos teóricos y prácticos de regresiones y de clasificadores.

En particular, a través de ejemplos simples usted será capaz de:

- Implementaren Python regresiones lineales, polinomiales y regresiones con penalización.
- Aplicar métodos disponibles por la librería Scikit-Learn para regresiones y clasificaciones.
- Experimentar y analizar el efecto de la complejidad de los modelos sobre el rendimiento de ellos.
- Visualizar modelos y resultados, utilizando Matplotlib.

## Trabajo a realizar

### Solución cerrada de regresión lineal

- Utilizando numpy (`np.random`), genere 100 datos que se encuentren entre 0 y 3. Este vector representará su variable independiente.
- Con la misma librería, genere los 100 datos de la variable dependiente de la siguiente forma  $y = 5 + 2x + u(-2,5)$ .
- Por medio de matplotlib (o seaborn), genere un gráfico de dispersión (scatter plot) entre la variable dependiente y la variable independiente.
- Estudie y aplique los métodos disponibles en *numpy.linalg* para ajustar una regresión lineal a través de su solución cerrada. ¿Cuál es el error cuadrático medio de la solución?
- Utilizando matplotlib (o seaborn), genere un gráfico que incluya la solución (recta óptima) y los datos utilizados.

### Regresión Polinomial

- De ahora en adelante, trabajaremos con una base de datos real, esta corresponde a una curva de luz (intensidad de luz de una estrella en función del tiempo). En primer lugar, lea el archivo *ts.csv* utilizando pandas. Trabaje con una muestra de 200 datos.
- Mediante matplotlib o seaborn, visualizar la curva de luz con un scatter plot.
- Genere una transformación polinomial de la variable independiente. Para esta transformación puede crear su propia función o puede utilizar el método disponible por Scikit-Learn.
- Ajuste regresiones polinomiales de diferentes grados.
- Grafique los modelos resultantes, ¿qué observa a medida que aumenta el grado del polinomio?

### Regresión con Penalización

- Utilizando la misma curva de luz *ts.csv* y polinomios de diferentes grados, experimente con regresión Ridge y Lasso.
- Obtenga los valores de los parámetros ¿Qué puede comentar al respecto? ¿Qué ocurre al aumentar el valor de alfa (penalización) en cada modelo?
- Grafique los modelos resultantes.

### Clasificación de estrellas variables

- Cargue la base de datos de estrellas variables *ts2.csv*. La columna *class* corresponde a la clase de cada estrella variable y el resto de las columnas corresponden sus descriptores.
- Genere un análisis descriptivo de ella y aplique las técnicas de procesamiento que sean necesarias.
- Separe la base de datos en dos conjuntos, uno para entrenar y otro para testear.
- Aplique 2 clasificadores de los vistos en el curso. Por cada uno de ellos, estudie el efecto que tienen los hiperparámetros sobre la predicción en el conjunto de entrenamiento y de testeo. Comente el efecto de cada parámetro.
- Estudie el clasificador KNN y use su implementación disponible en Scikit-Learn. Determine el número *k* (*n\_neighbors*) adecuado para este conjunto de datos.
- Compare el rendimiento de los clasificadores, utilizando matrices de confusión y métricas vistas en el curso.