

MINIPROYECTO 1- REGRESIÓN LINEAL MULTIPARAMETRICA Y POLINOMICA APLICADA A PROBLEMAS DE REGRESIÓN

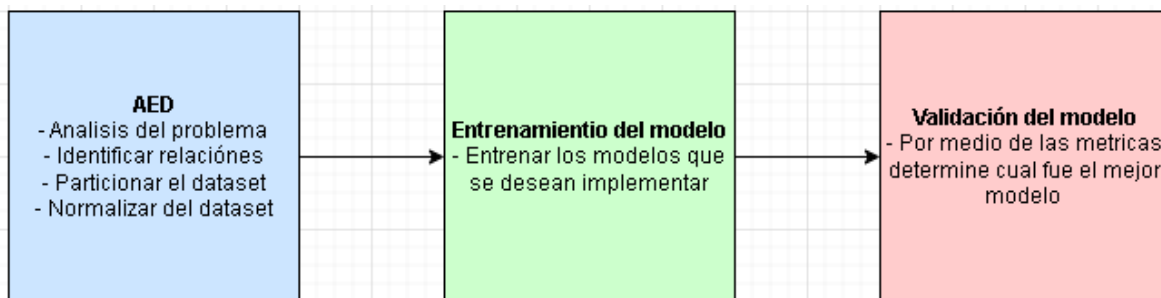
ENUNCIADO DEL EJERCICIO

Desarrolle dos modelos de machine learning, un regresor lineal multiparamétrico y un regresor polinómico los cuales tengan como objetivo resolver un problema de regresión que se podrá seleccionar de acuerdo a los dataset definidos (Ver subtítulo **DATASETS**). Deberá comparar los dos modelos obtenidos por medio de tres métricas de desempeño, el tiempo de entrenamiento y el tiempo de predicción. Además, deberá encontrar por lo menos tres relaciones entre sus variables por medio del análisis exploratorio de los datos (AED), por último plantee una hipótesis y compruébela con sus modelos obtenidos.

Recuerde que en el AED deberá encontrar la correlación de las variables y así mismo seleccionar como mínimo 3 características con las que desea trabajar bajo una sustentación sólida del porque seleccionaron dichas características.

ANOTACIONES

Se recomienda que tenga en cuenta los siguientes pasos:



Además podrá guiarse del documento visto en clase [Regresión Lineal Múltiple y Regresión Polinómica Ejemplo Práctico](#).

PUNTOS ADICIONALES

El estudiante que logre realizar por lo menos dos de los siguientes puntos obtendrá un 0.5 adicional en la nota de Miniproyectos.

- Implementar algún modelo adicional como el regresor lineal con regularización o un regresor bayesiano, entre otros.
- Implementar el regresor con una función propia sin la librería scikit learn.
- Implementar el regresor con el método del descenso del gradiente.
- Implementar LIME para interpretar el modelo seleccionado.
- Implementar un modelo de regresión distinto a los vistos en clase.

CALIFICACIÓN

El Miniproyecto 1 será calificado la clase 14 de la semana 7 donde se evaluará bajo los siguientes criterios:

- Sustentación (0.7)
- Informe en formato IEEE (0.3)
 - Introducción
 - Metodología
 - Desarrollo
 - Análisis y resultados
 - Conclusiones

DATASETS

En el listado que se encuentra a continuación se proponen los posibles datasets. Los estudiantes pueden proponer otros datasets diferentes, el cual debe ser avalado por el docente.

El número entre paréntesis indica la cantidad de grupos que pueden trabajar en cada dataset.

- **Airfoil Self-Noise Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>
- **Auto MPG Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Auto+MPG>
- **Energy efficiency Data Set (2)**(Cada grupo debe seleccionar una de las dos salidas disponibles y1= Heating Load, y2 = Cooling Load)
<https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>
- **Concrete Compressive Strength Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>
- **Yacht Hydrodynamics Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Yacht+Hydrodynamics>
- **Stock portfolio performance Data Set (3)** (Cada grupo debe escoger una salida de las seis posibles para realizar la regresión) Y1=Annual Return,Y2=Excess Return Y3=Systematic Risk,Y4=Total Risk ,Y5=Abs. Win Rate ,Y6=Rel. Win Rate
<https://archive.ics.uci.edu/ml/datasets/Stock+portfolio+performance>
- **Daily Demand Forecasting Orders Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Daily+Demand+Forecasting+Orders>
- **Real estate valuation data set Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set>
- **Electrical Grid Stability Simulated Data Data Set (2)** (En el enlace incluir el signo “+” al final)
<https://archive.ics.uci.edu/ml/datasets/Electrical+Grid+Stability+Simulated+Data>
- **Behavior of the urban traffic of the city of Sao Paulo in Brazil Data Set (2)**
<https://archive.ics.uci.edu/dataset/483/behavior+of+the+urban+traffic+of+the+city+of+sao+paulo+in+brazil>
- **QSAR fish toxicity Data Set (2)**
<https://archive.ics.uci.edu/ml/datasets/QSAR+fish+toxicity>