# Regresión de mínimos cuadrados

Por:

Ian Gabriel Cañas Fernández, 1092228

.

Profesor: Juan S. Pérez R.,

Asignatura: INL367L, Secc 01

## Resumen:

En el presente laboratorio se planea computar los pesos de una regresión lineal usando el algoritmo Least Squares, a partir de el cual se hará una estimación de la salida del dataset obtenido. A partir del comportamiento del dataset que asumimos como lineal se desarrollará una regresión polinómica para observar la relación entre ambos métodos y las posibles variaciones de este último.

## Ejercicios previos:

P6.1 Regresión lineal.

El objetivo principal de esta sección es el de computar los pesos de la regresión lineal a la simulación de un set de datos generado por el usuario. Este set sería segmentado en datos de entrenamiento y datos de prueba, siendo generado mediante la siguiente expresión:

Se puede observar de antemano que la expresión simularía la medición ruidosa de un par de características con correlación lineal. Los valores seleccionados para la expresión fueron:

P6.2 Predicción de las MPG.

Por otro lado, se espera la segmentación y predicción de un set de datos de Auto MPG Data Set. Se procurará computar los pesos de la regresión polinómica mediante el uso de Least Squares:

1. Considerando únicamente los dos *featucres* que arrojaron mejor desempeño.
2. Para 2 distintos sets de entrenamiento y pruebas.
3. Para 6 distintos grados de polinomios.

## E6.1 Naive Bayes; predicción de la supervivencia.

En la primera sección se genera, aleatoriza y segmenta el set de datos como se conoce de antemano mediante lo presentado a continuación:

Une image contenant texte

Description générée automatiquement

Más tarde, se genera una regresión lineal mediante la siguiente ecuación:

Continuando con la aplicación de los pesos obtenidos al set de pruebas de los datos.

Une image contenant texte

Description générée automatiquement

## E6.2 Naive Bayes; predicción de la supervivencia.

En la segunda sección, igualmente, empezamos aleatorizando los datos y seccionándolos en sets de entrenamiento y pruebas.

Une image contenant texte, intérieur, capture d’écran

Description générée automatiquement

Preparamos las matrices de características junto a su vector bias para la regresión. Para la primera prueba, mediante el uso de regresión a tres dimensiones, se hace selección de las dos mejores características. El método utilizado para esta sección fue la de recorrer las combinaciones posibles de las permutaciones mediante forward search.

Une image contenant texte

Description générée automatiquement

Aplicamos regresión lineal y revisa error mínimo cuadrado:

Une image contenant texte

Description générée automatiquement

Finalmente, reprocesando los datos, se preparan las matrices de características y bias. Se implementa regresión de mínimos cuadrados para varios grados polinómicos entre hasta el grado seis (6):

Une image contenant texte

Description générée automatiquement

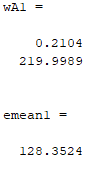
## R6 Resultados:

Para la primera sección se obtuvo los siguientes resultados de pesos y error cuadrado promedio para varias corridas del código:

Une image contenant texte

Description générée automatiquement Une image contenant texte

Description générée automatiquement Une image contenant texte

Description générée automatiquement 

Para la segunda sección se trabaja con la mejor combinación de características para conocer su eficiencia:

Une image contenant texte

Description générée automatiquement

Finalmente, se aplican las regresiones polinómicas:

Une image contenant texte

Description générée automatiquementUne image contenant texte

Description générée automatiquement

Une image contenant table

Description générée automatiquementUne image contenant table

Description générée automatiquement

Une image contenant table

Description générée automatiquementUne image contenant table

Description générée automatiquement

## A.5 Análisis:

En base a los resultados obtenidos mediante la regresión, se observó que los resultados de pesos obtenidos mediante la primera sección se veían directamente relacionados al error cuadrado promedio. Este error tiende más a parecer menor según los pesos de la predicción se asemejen a los computados anteriormente en la generación de los datos. Por lo que este método tiende a no ser muy alterado ante el ruido siempre y cuando este tenga su media en el cero.

Luego, comprobando lo obtenido con las dos mejores características y con varios grados de polinomios, se observó que los grados de polinomios afinan su eficiencia según se aumente dicho grado. Sin embargo, si se desea trabajar con pocas características representativas, es conveniente el uso de la mejor combinación, tomando en cuenta que trabajar con muchas características para este método no es computacionalmente tan pesado.