

MODELOS BASADOS EN REGRESIÓN LINEAL PARA PREDECIR EL PRECIO DE LAS VIVIENDAS.

Fue solicitada la implementación de dos modelos basados en regresión lineal para predecir el precio de las viviendas. Esto con el fin de saber cuál sería un buen precio de venta.

Se realizó el planteamiento de ambos modelos de Machine Learning con los datos de entrada (datos normalizados y no normalizados) y procedimos a entrenar ambos modelos con el 60% de los datos, luego utilizamos el 40% restante para probar la exactitud del modelo entrenado. Cabe aclarar que los datos fueron tomados al azar dentro de la lista para ambos propósitos.

Para los dos modelos se trabajó con **10000** iteraciones y un alfa de **0.0000000001**. Los resultados de la hipótesis y el MAPE (Mean Absolute Percentage Error) de ambos modelos fueron los siguientes:

Datos normalizados:

Hipótesis: $h(x) = (-5.894312815613162 \text{ e-}23) + (6.350446323769998 \text{ e-}08) x_1 + (3.3234545304135806 \text{ e-}08) x_2$

$$\theta_0 = -5.894312815613162 \text{ e} - 23$$

$$\theta_1 = 6.350446323769998 \text{ e} - 08$$

$$\theta_2 = 3.3234545304135806 \text{ e} - 08$$

MAPE: 99.99999771798367 %

Datos no normalizados:

Hipótesis: $h(x) = (0.08870216000560295) + (167.23910926085873) x_1 + (0.27678573947258395) x_2$

$$\theta_0 = 0.08870216000560295$$

$$\theta_1 = 167.23910926085873$$

$$\theta_2 = 0.27678573947258395$$

MAPE: 11.933795770424648 %

Análisis: podemos observar que existe un MAPE muy elevado, casi del 100%, a comparación del MAPE de los datos no normalizados, esto es debido a que el MAPE funciona mejor cuando no hay extremos en los datos (incluyendo los ceros).

Con valores cercanos a ceros o ceros, el MAPE proporcionara una imagen distorsionada de error general cuando este es promediado. Para el caso de tener estos valores lo mejor es usar el SMAPE (Symmetric mean absolute percentage error) que nos brinda una mejor medida.

MAPE (Mean Absolute Percentage Error):

Ventajas:

- MAPE no es compleja de implementar.
- No es dependiente de escala y puede aplicarse sin complejidad a altas y bajas cantidades de valores.

Desventaja:

- Con valores cercanos a ceros o ceros el MAPE da como resultado una imagen distorsionada de error general.

Nombre: Kevin A. Vega Hernández.

CD. T00038080.