Universidad del Valle de Guatemala

Minería de Datos

24 de abril de 2020

Antonio Reyes Acevedo

Dieter de Wit

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Hoja de Trabajo 7

Redes Neuronales Artificiales y Máquinas Vectoriales de Soporte

## Explicación de Redes Neuronales

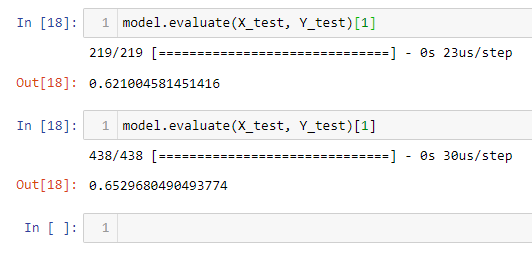
Las variables de entrada utilizadas para la Red Neuronal son las mismas variables que se utilizaron para poder definir la clasificación de rangos de precio en HT-7 Parte 1, tanto los datos de entrada y la columna con las clasificaciones (los datos que se quieren predecir) se encuentran en un archivo .csv definido en el notebook. Esta data se convierte en arrays y se separan todos los datos como la variable X y el dato a predecir en Y.

Para preparar la información para red neuronal, se escala el conjunto de datos para poder facilitar el entrenamiento de la red neuronal y se separan la base de datos en data de entrenamiento y testing. Esta información se utilizará en el modelo proveído por Keras que consiste de dos capas de 32 neuronas y la capa de salida cuyo número de neuronas dependerá del número de categorías de casas.

Ya que se tiene el modelo y los datos, se utiliza la función “fit” para ajustar los parámetros a los datos donde se especifica los datos a entrenar, los datos de validación y la cantidad de iteraciones que se harán para entrenar el modelo.

model.evaluate(X\_test, Y\_test)[1] determina la exactitud.

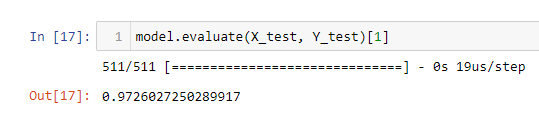
## Resultado de Redes Neuronales con primera lista de rangos

El primer intento de predicción de las categorías de las casas dio como resultado un 62.10% de exactitud, pero al aumentar el número de iteraciones y número de datos de entrenamiento, aumento de 65.30% como se ve en la imagen adjunta. 

Se intentó aumentar el número de iteraciones, pero se generaba un error en el cual, después de cierta cantidad de iteraciones, los valores de **loss, accuracy** y **var\_loss** solo mostraban cero.

## Resultado de Redes Neuronales con segunda lista de rangos

Debido a esto, se realizó otra predicción de las categorías en la que, en base a los mismos datos, se iba a predecir si la casa pertenecía a un rango específico entre 0-200K, 200K – 400K, 400K-600K y 600K-800K. Con esta lista de rango se obtuvo una predicción con 97.26% de exactitud.



En base a los resultados de los modelos previos, puede verse que los resultado la predicción de las redes neuronales son más eficientes pero requiere más preparación y tiempo ya que un aspecto muy importante de este modelo es tener una mucha información de entrenamiento y repetición para optimizar los resultados.

## Comparación de SVM y Redes Neuronales

Al evaluar los resultados de certeza de los modelos de SVM, puede verse que las Redes Neuronales tuvieron un mejor porcentaje en la segunda lista de rangos, pero se tuvo un mal rendimiento en la primera lista de rangos (este bajo rendimiento se debe al problema en las iteraciones, que volvía los valores 0 si se modificaba el numero de iteraciones). En base a estos resultados puede concluirse que los modelos SVM y Redes Neuronales has probado ser más eficientes que los modelos vistos en las previas hojas de trabajo, pero que la eficiencia del modelo de Redes Neuronales depende directamente de la cantidad de información de entrenamiento y de el tiempo de entrenamiento del modelo.