

Ex08 Análisis de Sentimientos

1. Describe las características del conjunto de datos `IMDB Dataset`

La entrada es una matriz de 50000x2. La primera columna son frases todas en inglés, algunas con caracteres propios de HTML la segunda columna es la connotación de la frase ya sea positiva o negativa.

2. Indica el tipo de procesamiento que se realiza al texto.

Se limpia cada oración de la siguiente manera:

- a) Se quita cualquier carácter que no sea parte del alfabeto.
- b) Se quitan caracteres aislados.
- c) se quitan espacios en blanco que haya de más.

3. Indica que porcentaje de datos utilizaste para:

- a. Entrenar : 80%
- b. Probar: 20%

4. ¿Explica qué es `GloVe embeddings`?

“Gloval Vectors” son vectores para representar una palabra. Se basa en modelos que producen “Word embedding”. No soy un experto en el área pero en base a lo que leí en un blog(<http://techscouter.blogspot.com/2017/10/word-vectorization.html>) al parecer hay grupos de palabras que están muy relacionadas y forman grupos. Mientras mas central o importante sea una palabra en el grupo tendrá un valor más alto o peso. Cada elemento dentro del vector es uno de esos pesos. En esta base de datos los valores están normalizados de -1 a 1.

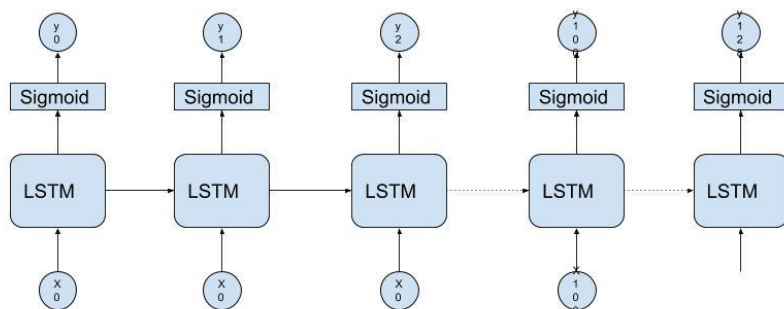
5. ¿Qué es una embedding `matrix`?

Es una matriz en la que cada fila es un GloVe, y el índice de cada fila es el índice de cada palabra del vocabulario.

6. Describe la arquitectura de tu modelo de red neuronal recurrente.

- a. Longitud de secuencia de entrada: 100
- b. Longitud de secuencia de salida: 128 para Y pero al final solo sale una respuesta del de 0 o 1 si dependiendo el sentimiento.
- c. Número de capas y neuronas por capa: Una capa Hidden de 128 neuronas.
- d. ¿Qué funciones de activación utilizas? Sigmoid pero internamente dentro de la LSMT se usa también Tanh.
- e. ¿Qué función utilizas para la predicción? Model.predict pero no se usa en este código aquí directamente se evalúe todo el batch pruebas con evaluate.

7. Dibuja la estructura de tu red neuronal recurrente (LSTM).

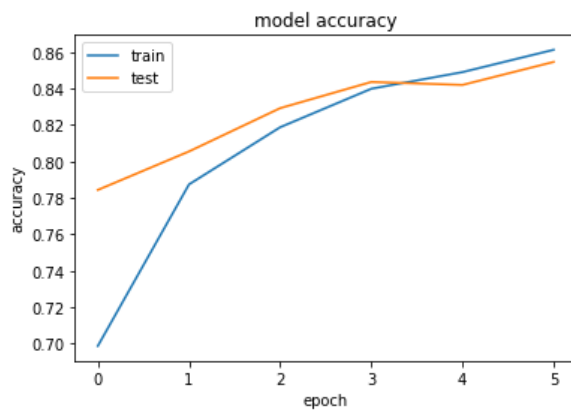


8. ¿Cuál fue la exactitud del modelo?

0.8547

9. Incluye las gráficas:

a. Model accuracy



b. Model los

