



Tarea 3.- Segundo Parcial

Modelos de clasificación Múltiple

Instrucciones:

Para esta tarea, desarrollarás el código de **3 clasificadores de texto basado en redes neuronales con Word Embeddings** (Una DNN, una CNN y una LSTM) y realizarás una comparación entre los resultados de los mismos.

En esta tarea encontrarás un archivo adjunto (.zip), el cuál contiene 4 tipos de textos, cada uno describe alguna entidad diferente (Personas, Objetos, Lugares, y Animales). Estos textos fueron extraídos directamente de Wikipedia (Y recortados un poco para no ser tan extensos).

- 1) Deberás descomprimir el archivo con las carpetas de textos y agregar una carpeta nueva con algún tema de tu preferencia, por ejemplo, libros, o películas, y agregar textos de similar extensión a los otros ejemplos, ya que estos ejemplos que agregues, serán una nueva categoría para clasificar en tu ejercicio.
- 2) Deberás desarrollar el código necesario para ejecutar TODO el pipeline para crear, probar y evaluar los 3 tipos de modelos de clasificación (DNN, CNN y LSTM), incluyendo:
 - Lectura de información o Corpus
 - Preprocesamiento del texto
 - Creación de las matrices de entradas (X) y de salidas (Y)
 - Creación de la matriz de embeddings para el diccionario de tu corpus
 - Definición de los modelos DNN CNN y LSTM MULTICLASE (Con los hiperparámetros que consideres más adecuados)
 - Entrenamiento de los modelos
 - Muestra gráfica del accuracy y pérdida de cada uno de los modelos
 - Matriz de confusión para cada uno de los modelos.

NOTA: TODO el código deberá de contar con uso de buenas prácticas, tales como, nombres apropiados para métodos y variables, uso correcto de DocStrings y comentarios, no tener redundancia de código, etc.

- 3) Deberás de programar todo lo anterior y realizar un reporte con los siguientes segmentos:
 - A) BREVE Introducción: ¿Qué categorías se van a clasificar?, ¿Qué se realizó en el código? (Puntos importantes como la selección de hiperparámetros o ¿Por qué se eligió cierto tipo de preprocesamiento?
 - B) Desarrollo: Aquí deberás de mostrar capturas de pantalla de TODO el código (Explicando brevemente qué haces en cada celda) y sus respectivas salidas, desde la lectura del corpus hasta las matrices de confusión de cada modelo y la comparativa de las mismas
 - C) Conclusión: Indicar ¿Qué observas en la comparación de los modelos?, ¿Cómo afectan los hiperparámetros de cada tipo de red neuronal?, ¿Qué se complicó en el clasificador?, ¿Cómo se podría mejorar el modelo?, entre otras cosas que se consideren importantes.



4) En esta tarea deberás de entregar el reporte descrito arriba en formato PDF, además del notebook donde hayas realizado todo el desarrollo del entrenamiento y validación de tus modelos, y el archivo .zip donde se encuentren todos los textos que hayas utilizado para la clasificación.

TODOS los integrantes del equipo (De 1 a 3 personas) deberán subir TODOS los archivos de la tarea solicitados arriba ANTES de la fecha final de entrega, y el nombre de TODOS los integrantes debe estar presente en el reporte en PDF. De lo contrario, no se tomará en cuenta la evaluación.