

## Redes Neuronales para Lenguaje Natural

Noviembre 2023

### Consideraciones generales

- i) La prueba es sin material escrito.
- ii) Escriba nombre y C.I. en todas las hojas.
- iii) Numere todas las hojas.
- iv) En la primera hoja, indique el total de hojas.
- v) Comience cada ejercicio en una hoja nueva.
- vi) Utilice las hojas de un solo lado.
- vii) Entregue los ejercicios en orden.

### Ejercicio 1

Considere la siguiente tarea de PLN: Dado un texto, se desea obtener los nombres de personas que aparecen en el texto y sus profesiones. Se cuenta con un corpus de 5000 textos con los nombres propios y profesiones anotados, como el siguiente:

*[PER José Rodríguez] vive en Canelones, y su esposa [PER Amalia] [PROF atiende un kiosco] que queda al lado de la casa de tejas de la esquina. Él trabajó toda su vida de [PROF mecánico] en un taller que queda en el centro.*

En esta tarea, se desea poder ingresar un nuevo texto cualquiera en español, y obtener como resultado las marcas PER y PROF, como las anteriores, en los lugares apropiados del texto. No se solicita indicar qué profesión corresponde a cada persona.

- a) Indique por lo menos dos arquitecturas de redes neuronales de las vistas en el curso que serían apropiadas para esta tarea.
- b) Diseñe una red para resolver esta tarea indicando entradas, salidas, capas utilizadas y funciones de activación.
- c) Describa cómo realizaría el entrenamiento de la red de la parte anterior, incluyendo cómo partiría los datos y cómo elegiría los mejores hiperparámetros.

### Ejercicio 2

Responda las siguientes preguntas sobre el modelo de red neuronal Transformer.

- a) Describa brevemente en qué consiste el modelo Transformer y mencione dos ejemplos de uso.
- b) ¿Qué pesos son ajustados durante el entrenamiento en un bloque Transformer en lo que respecta al multihead self-attention?
- c) ¿Cómo es considerado el orden de las palabras en el modelo Transformer?

### Ejercicio 3

Elija solo una respuesta correcta para las siguientes preguntas de múltiple opción:

- 1) Los vectores de palabras (word embeddings) son:
  - a) Una arquitectura de red neuronal para resolver problemas secuenciales.
  - b) Un método para crear vectores de reales densos que aproximan la semántica de las palabras.
  - c) Una función de activación que se basa en tomar el coseno del ángulo entre dos vectores.
  - d) Una forma de representar textos donde cada palabra puede valer 0 o 1 en un vector del tamaño del vocabulario.
- 2) Una capa convolucional unidimensional:
  - a) Opera sobre las dependencias de largo alcance de la entrada.
  - b) Opera sobre colocaciones locales de la entrada.
  - c) Obtiene una representación de la entrada mediante un mecanismo de atención.
  - d) Retorna una salida de 1 dimensión que permite realizar clasificaciones binarias.

- 3) La función de activación softmax:
  - a) Permite transformar las salidas de una capa en una distribución de probabilidad
  - b) Devuelve valores entre  $-\infty$  y  $+\infty$
  - c) No puede utilizarse porque no es derivable en todos sus puntos
  - d) Se utiliza principalmente cuando el vocabulario de entrada es diferente que el vocabulario de salida
  
- 4) Las redes tipo LSTM:
  - a) Son el único tipo existente de red recurrente
  - b) Se basan en el mecanismo de auto-atención para procesar la entrada
  - c) Procesan un elemento por vez, al que combinan con los estados ocultos que venían procesando, para generar la salida
  - d) Realizan una convolución de la entrada a la que luego se le aplica una capa de pooling
  
- 5) ¿Qué es el learning rate?
  - a) La cantidad de capas de una red.
  - b) Una función no lineal que se utiliza a la salida de una capa.
  - c) La magnitud por la que se modifican los pesos según el gradiente durante el entrenamiento.
  - d) Un número que indica la cantidad de veces que se itera completamente sobre el corpus de entrenamiento al entrenar.
  
- 6) ¿En qué consiste el concepto de fine-tuning?
  - a) La capacidad de los modelos para reducir la función de pérdida sobre el corpus de entrenamiento.
  - b) La adaptación de un modelo pre-entrenado utilizando un conjunto más pequeño de datos distinto del original.
  - c) El ajuste de los hiperparámetros para llegar a un modelo que se adapte más a los datos.
  - d) El efecto cuando un modelo tiene alto desempeño sobre los datos de entrenamiento, pero esto no se ve reflejado en los datos de test.

#### Ejercicio 4

Indique si las siguientes afirmaciones son verdaderas o falsas. Justifique en ambos casos.

- a) Un Transformer es un modelo de red neuronal recurrente con atención.
- b) BERT es una alternativa a GPT para generar secuencias.
- c) El early stopping es una forma de controlar que el modelo no se sobreajuste durante el entrenamiento.
- d) El algoritmo BPE se utiliza para construir word embeddings.
- e) La técnica de few-shot prompting implica presentar unos pocos ejemplos de cómo se resuelve una tarea en un gran modelo de lenguaje.
- f) Los positional embeddings son representaciones para indicar la posición de cada palabra en una secuencia y son utilizados en los modelos de redes recurrentes debido a que no consideran el orden de las palabras.
- g) El dropout es una técnica para tratar las palabras fuera de vocabulario en un texto.
- h) LORA (Low-Rank Adaptation) es una técnica de reducción de dimensionalidad que puede usarse para visualizar conjuntos de embeddings.