

## Ejercicio 2 - Repaso Teoría

### 1. ¿Cuál es la principal innovación de la arquitectura Transformer?

La principal innovación de la arquitectura Transformer es eliminar las capas recursivas con la finalidad de beneficiar a los mecanismos de autoatención. Después de eliminar las capas, el entrenamiento debería de ser más paralelizado y se podría reducir bastante el tiempo. También hay que destacar que la arquitectura es muy eficaz en tareas de traducción de lenguaje natural.

### 2. ¿Cómo funciona el mecanismo de atención del scaled dot-product?

El mecanismo scaled dot-product es un mecanismo que se basa en calcular el producto escalar entre una query y varias keys para cada posición en una secuencia. Después cuando ya se tiene el producto se le aplica una función softmax para obtener los pesos. Ya de ultimo se utiliza estos pesos para ponderar los valores correspondientes a las claves y obtener la salida de atención para cada posición.

### 3. ¿Por qué se utiliza la atención de múltiples cabezales en Transformer?

Principalmente se utiliza para que el modelo adquiriera una visión mas paralela y pueda manejar diferentes aspectos de la información. Cada cabezal proyecta las queries, keys y valores a subespacios diferentes, lo que permite que el modelo pueda aprender representaciones más diversas y distintas en relación con los datos.

### 4. ¿Cómo se incorporan los positional encodings en el modelo Transformer?

Estas se incorporan proporcionándole información al modelo sobre la posición relativa de las palabras en la secuencia, supliendo esa necesidad de la arquitectura de no tener una noción incorporada de orden secuencial.

### 5. ¿Cuáles son algunas aplicaciones de la arquitectura Transformer más allá de la machine translation?

Otras aplicaciones podrían ser:

- El análisis de sintaxis (En el que más se destaca)
- Resumen abstractivo
- Comprensión de lectura
- Entailment textual
- Representaciones de oraciones independientes de la tarea.