TPLN: from embeddings to transformers

Puntos totales 7/10

For each of the questions, write your answer here along with a brief justification for it. This justification should be the minimum required for the instructor, who is also knowledgeable in the subject, to conclude that you are capable of defending your answer and that it is not a result of chance or other collateral effects. Remember that this form can be filled out until the date and time specified on the course website; responses submitted after this will be disregarded.

Se ha registrado el correo del encuestado (**dar33@gcloud.ua.es**) al enviar este formulario.

✗ The skip-gram method as studied in the Jurafsky and Martin's book allows for a concrete tuple (w,c) to appear in both the set of positive and negative samples. 0/1

- ○ True
- ◉ False ✗

Respuesta correcta

- ◉ True

Justification for the answer to the previous question

En el método skip-gram, según lo estudiado en el libro de Jurafsky y Martin, una tupla concreta (w, c) no puede aparecer tanto en el conjunto de muestras positivas como en el negativo.

✓ The skip-gram method as studied in the book provides an embedding for each word represented by the corresponding vector in the W matrix at the end of training. The matrix C is always discarded as it contains random vectors due to the stochastic nature of the gradient descent algorithm used during training. 1/1

- ○ True
- ◉ False ✓

Justification for the answer to the previous question

En el método skip-gram, la matriz W contiene los embeddings de las palabras después del entrenamiento, y es importante en el modelo. Sin embargo, la matriz C no se descarta simplemente por contener "vectores aleatorios". Tanto la matriz W como la matriz C se actualizan durante el entrenamiento utilizando el descenso de gradiente estocástico

✓ It has been proved that, without additional adjustments, word2vec algorithms usually amplify the gender biases in the training set, but not the racial biases. 1/1

- ○ True
- ◉ False ✓

Justification for the answer to the previous question

Los algoritmos de word2vec, como el método skip-gram, no están limitados a amplificar solo los sesgos de género en el conjunto de entrenamiento también pueden amplificar sesgos raciales y otros tipos de prejuicios.

✓ The embedding obtained with the skip-gram method for the word "Mars" will be different for the cases in which the word refers to the Roman god of war or to the closest planet to the Earth. 1/1

- ○ True
- ◉ False ✓

Justification for the answer to the previous question

El método skip-gram de word2vec, tal como se utiliza comúnmente, no distingue entre diferentes significados de una palabra.

✓ A language model implemented via feedforward neural networks will make better predictions if every input word is represented with a one-hot vector than if they are represented with a word embedding precomputed with the skip-gram method. 1/1

- ○ True
- ◉ False ✓

Justification for the answer to the previous question

Los modelos de lenguaje implementados mediante redes neuronales alimentadas hacia adelante generalmente hacen mejores predicciones cuando cada palabra de entrada está representada por un embedding de palabras precalculado con el método skip-gram, en lugar de utilizar vectores one-hot.

✗ The masking mechanism used in the decoder of a transformer is intended to prevent the model from accessing future information during training, although ignoring the future is not necessary during inference. 0/1

- (●) True ✗
- ( ) False

Respuesta correcta

- (●) False

Justification for the answer to the previous question

En los decodificadores de los transformers, el mecanismo de enmascaramiento se utiliza para prevenir que el modelo acceda a información futura durante el entrenamiento.

✓ Skip connections, also known as residual connections, in the transformer architecture, allow the gradient to flow directly through the network, mitigating the vanishing gradient problem. 1/1

- (●) True ✓
- ( ) False

Justification for the answer to the previous question

Las conexiones residuales (skip connections) en la arquitectura de los transformers permiten que el gradiente fluya directamente a través de la red

✓ The positional encoding in transformers is added to the input embeddings to provide a sense of word order. 1/1

- (●) True ✓
- ( ) False

Justification for the answer to the previous question

En los transformers, el codificado posicional se añade a los embeddings de entrada para proporcionar una noción del orden de las palabras.

✗ The book mentions that layer normalization in the transformer architecture is applied before the multi-head attention and feed-forward layers because this is how it was implemented in the origina paper introducing the model, but nowadays it is more common to apply it after these layers. 0/1

- (●) True ✗
- ( ) False

Respuesta correcta

- (●) False

Justification for the answer to the previous question

En el libro se menciona que en la arquitectura original del transformer la normalización de capas se aplicaba antes de las capas de atención multi-cabeza y de feed-forward, como se implementó en el artículo original que introdujo el modelo.

✓ In a transformer model, the number of heads in the multi-head attention mechanism directly influences the dimensionality of the embeddings outputted after each layer. 1/1

- ○ True
- ◉ False ✓

Justification for the answer to the previous question

El número de cabezas en el mecanismo de atención multi-cabeza no influye directamente en la dimensionalidad de los embeddings resultantes después de cada capa.

Finally, write down one or more questions you still have after studying the materials or taking this test. Try to be as precise as possible. This last question does not count towards the assessment, but please answer it for the effective functioning of the flipped classroom system. The instructor will not mention your name if your question is discussed during the class. "No questions, thank you" or similar are not valid answers. *

No questions, thank you