

NLP y Grafos Probabilísticos

2 de agosto 2021

Indicaciones

Para la resolución de la práctica se espera de los alumnos que no solo sean capaces de aplicar el código para la resolución de los ejercicios, si no que también expliquen los procesos que realizan y muestren el conocimiento que han adquirido a la hora de explicar los pasos que dan y porque hacen las cosas de una determinada manera.

Se entiende que son capaces de aplicar los conceptos como train/test, etc. aprendidos en los anteriores bloques, así como el análisis de los resultados en función de los datos que se tienen y ser capaces de identificar posibles sesgos en los datos.

Por ello, han de tener en cuenta que la evaluación de esta práctica no solo se realizará sobre el código que escriban y los resultados que obtengan.

- Una práctica sin explicaciones pero con todo el código correcto en ningún caso se evaluará con la máxima nota.
- Una práctica que en algún caso no alcance un benchmark o unos resultados objetivo o cuyo código tenga pequeños errores, si los razonamientos y las explicaciones apoyan el código podrá alcanzar la nota máxima.

1. Procesamiento del Lenguaje Natural

1.5p – 1 A partir del dataset FinancialPhraseBank utilizando unicamente las frases financieras con un 100% de coincidencia en la etiqueta realizar las siguientes tareas:

- Bag of words
- TF-IDF
- Word-to-vect

0.5p - 2 Visualizar los resultados obtenidos en el ejercicio 1 mediante alguna técnica de reducción dimensional si fuera necesario.

5p - 3 Mediante el dataset Financial PhraseBank (realizar para el dataset completo y solo para las frases 100% agree) construir un clasificador de sentimientos con las siguientes técnicas. Comparar sobre test los resultados de vuestro modelo con algún modelo de análisis de sentimiento ya implementado como nltk y etc:

Se evaluarán los resultados en función de la distancia al benchmark objetivo 0.86 acc (alldataset) y 0.98 acc(allagree), la complejidad de la solución (embeddings utilizados y entrenados, etc) y la explicación del proceso.

- Redes recurrentes con capas de atención.
- Redes recurrentes con convolucionales.
- Transformers.

2. Modelos de Grafos Probabilísticos

1.5p - 1 Utilizar el dataset african_crises.csv para construir una red bayesiana y realizar las siguientes tareas.

- Realizar la tarea de clasificación sobre la variable banking_crisis sin utilizar la variable systemic_crisis. No se evaluará sobre la precisión del modelo ya que es una dataset desbalanceado. Se evaluará sobre las explicaciones e interpretación de los resultados, arquitectura y construcción del modelo.
- Realizar inferencias dadas las evidencias que consideréis para extraer información y explicar los resultados obtenidos.

1.5p - 2 Utilizar cualquier activo que cotice en un índice de los siguientes [SP500. STOXX50, IBEX35] o el mismo índice detectar los regímenes utilizando una o varias variables de observación.