




TALLER DE PROCESAMIENTO DE SEÑALES

1er Cuatrimestre 2026 - Trabajo Práctico Nº 8 - Modelos de Lenguaje y Sistema de Recomendación

1. Se desea evaluar el funcionamiento de un modelo de lenguaje. Para ello, se utilizará la relación semántica entre países y sus capitales.

(a) *Modelo de lenguaje:*

- a) Descargar las representaciones pre-entrenadas del modelo de lenguaje *FastText* en idioma inglés de dimensión 300. : El modelo preentrenado puede descargarse en <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.en.300.bin.gz>
- b) Utilizando `load_model` (`fasttext`), cargar las representaciones pre-entrenadas. : En este TP solo utilizaremos `FastText` como modelo de lenguaje, queda prohibido utilizar sus métodos.

(b) *Analogías:* Combinar los países y las capitales del archivo `country-list.csv` para implementar una función que conteste a la pregunta *A es a B como C es a ...* dando 3 opciones posibles. Para ello, combine las vectorizaciones de A, B y C y devuelva las 3 palabras más cercanas de la base de datos que acaba de definir. : Ej. *France* es a *Paris* como *England* es a ... (London).

(c) *Representación gráfica:* Utilizando las vectorizaciones de las palabras de la base de datos, entrenar un algoritmo PCA que reduzca la dimensión a 2 los nombres de países y capitales del archivo `country-list.csv`, utilizando `decomposition.PCA` (`sklearn`). Graficar las representaciones reducidas de *Italy*, *Rome*, *France*, *Paris*, *Germany* y *Berlin* uniendo los países con sus capitales.

2. Se desea diseñar el sistema de recomendación de materias de FIUBA.

(a) *Análisis Teórico:* Hallar, analíticamente, el gradiente del *riesgo empírico regularizado* correspondiente a un filtro colaborativo. Expresar los resultados vectorizados.

(b) *Preprocesamiento:*

- a) Asignar 1 a las calificaciones positivas, -1 a las negativas y 0 a los campos sin completar.
- b) Inventar un estudiante con 15 materias calificadas. Explicar que tipos de materias le gustan al estudiante ficticio y sea consistente.

(c) *Entrenamiento:*

- a) Implementar, utilizando *gradiente descendente*, un filtro colaborativo para un espacio latente de dimensión 5.
- b) Repetir el entrenamiento eligiendo valores razonables para el *learning rate* α y el multiplicador λ . Se debe alcanzar la convergencia y poseer un valor de $\lambda > 0$ tal que el filtro no recomiende las mismas materias para todos los usuarios.

(d) *Rating:*

- a) Crear un *rating* ponderando la salida del filtro colaborativo y la calificación media de las películas utilizando un hiperparámetro p . Elegir un valor de p tal que no le recomiende a todos los estudiantes las mismas materias, pero que las recomendaciones difieran del caso $p = 1.0$.
- b) Recomendar al estudiante ficticio las 5 materias **no calificadas** con más alto *rating*. Extraer conclusiones.