

Laboratorio 2

Descripción y objetivo

El laboratorio está diseñado para familiarizar al estudiante con el ajuste de modelos de recomendación basados en factorización

Material

- Python 3.5
- Framework surprise
- *Dataset* Movielens 10k disponible en Sicua
- Notebook Laboratorio 2.ipynb

Metodología

Realice el laboratorio en los grupos previstos para el trabajo práctico del curso. Se realiza una entrega por grupo.

Entregable

La entrega de resultados del laboratorio consta de dos partes: El software desarrollado y un informe de laboratorio.

Realice un informe donde documente su desarrollo y avance de cada uno de los puntos del laboratorio. Inicialmente muestre, mediante imágenes de pantalla, su logro de los objetivos de cada punto.

En los puntos en los cuales se espera que usted explore y experimente, realice una pequeña descripción de lo que encuentra y obtiene.

Documente los resultados obtenidos en los desarrollos realizados y haga un breve análisis sobre los resultados.

Formato y hora de entrega

Para la presentación del informe de laboratorio utilice la plantilla de documentos disponible en Sicua+.

Realice su entrega del laboratorio de la siguiente forma:

- Archivo en formato **zip**, nombrado de la siguiente forma: **Lab2_NN_login1_login2_login3.zip**, donde NN es el número del grupo y luego se encuentran los login uniandes de los integrantes del grupo.
- Contenido del archivo:
 - Notebook jupyter desarrollado en el laboratorio
 - Documento de informe en formato **pdf**, nombrado de la siguiente forma:
Lab2_NN_login1_login2_login3.pdf

Fecha límite de entrega: **Fecha programada de laboratorio, a la hora de finalización de clase, en Sicua+.**

Desarrollo

0. Alistamiento de datos y ambiente de desarrollo

Descargue los datos previstos para el laboratorio

- a) Si el computador en el que va a trabajar no tiene Anaconda, descargue Miniconda (Python 3.6) para su plataforma, desde <https://conda.io/miniconda.html>

- b) Si no lo tiene, instale jupyter notebook en su máquina (comando: conda install jupyter)
- c) Importe el ambiente virtual suministrado en el yml de sicuaplus <https://conda.io/docs/user-guide/tasks/manage-environments.html>
- d) Ejecute jupyter notebook, abra el notebook enviado por sicuaplus utilizando como kernel el ambiente virtual suministrado

1. Exploración de datasets de entrenamiento, validación y test

- a. Siga el link de documentación del framework presentado en el notebook
- b. Explore ejecutando el notebook cómo se distribuyó el dataset original
- c. Documente sus hallazgos

2. Entrenamiento modelo SVD con Bias

- a) Ejecute el modelo SVD con Bias presente en el Notebook
- b) Documente los resultados obtenidos al correr el modelo, ¿qué efecto sobre los parámetros obtenidos observa?, ¿qué puede decir sobre la matriz original de ratings de entrenamiento? ¿Qué puede decir sobre el error absoluto en las coordenadas conocidas?

3. Evaluación de predicciones

- a. Ejecute la generación de predicciones para los datasets de entrenamiento, validación y prueba
- b. De acuerdo a los resultados obtenidos, ¿Qué puede decir del modelo SVD entrenado con los hiper-parámetros suministrados?

4. Selección de hiper-parámetros

- a. Aumente el número k de dimensiones de vector y documente los resultados. Evidencie el efecto que tiene esta variación en los parámetros obtenidos y en los errores (cree una gráfica donde muestre el efecto que tiene k sobre los errores)
- b. Para un k fijo, aumente la regularización para diferentes valores [0.001,0.1,0.5,1,10]. Evidencie el efecto que tiene esta variación en los parámetros obtenidos y en los errores.

5. Conclusiones

Escriba en máximo 3 párrafos las conclusiones que tiene sobre este laboratorio