

Practica 8: Sistemas de recomendación

Objetivo

El objetivo es implementar un sistema de recomendación de películas basado en las valoraciones de los usuarios.

Estudio previo (es necesario prepararlo por escrito, antes de acudir a la práctica)

1. Repasa las transparencias de clase y estudia las funciones auxiliares proporcionadas para esta práctica. En particular, necesitas revisar y comprender bien los gradientes de la función de coste vista en clase. Se recomienda, hacerlo en papel.

Desarrollo de la práctica

Copia a tu directorio de trabajo los ficheros proporcionados, y comprueba que funcionan correctamente en Matlab.

1. La práctica utiliza la base de datos MovieLens 100k Dataset from GroupLens Research. La base de datos consiste en las calificaciones de 943 usuarios a 1682 películas. Las recomendaciones están en el archivo `p8_movies.mat`. Al cargar este archivo aparece una matriz Y de recomendaciones ($\text{num_películas} \times \text{num_usuarios}$) y una matriz R del mismo tamaño cuya componente $R(I,J)$ indica si la película I ha sido calificada por el usuario J .
Siguiendo la notación usada en clase, los parámetros de cada usuario y las características de cada película se denotan con
 - $\text{Theta}(\text{num_usuarios}, \text{num_características})$, $\text{Theta}(J,:)$ es un vector que contiene los parámetros del usuario J
 - $X(\text{num_películas}, \text{num_características})$, $X(I,:)$ es un vector que contiene las características de la película I

En la práctica el número de características es de 100, por lo que los vectores anteriores tienen dimensión 100.

2. El guión de la práctica esta contenido en el archivo `p8.m`, que además os proporciona una verificación de cada uno de los pasos. **Deberéis completar la función `cofiCostFunc.m`**. Antes de avanzar en cada paso, aseguraos de que el resultado que se obtiene es igual al que proporciona el script `p8.m`. Los pasos son:

1. Leer los datos
2. Implementar la función de coste (completar `cofiCostFunc.m`)
3. Añadirle el gradiente
4. Añadirle la regularización
5. Comprobar la función de coste completa
6. Introducir tus calificaciones
7. Entrenar el recomendador
8. Predecir tus propios ratings y generar recomendaciones para ti mismo

Se recomienda hacer una primera implementación con bucles y después, como trabajo opcional, optimizarla para matlab con cálculo matricial. Si no se hace la implementación matricial, la optimización irá mucho más lenta.

Para la optimización de X y Θ , se va a usar la función `fmincg` (similar a `minFunc`, utilizada en prácticas anteriores). El desenrollado de las matrices X y Θ en `cofiCostFunc.m` es necesario para poder utilizar este optimizador.

A entregar (en Moodle, dentro de un fichero .zip)

- Programa `p8.m`, junto con las funciones auxiliares que hayas programado, que vaya mostrando por pantalla los resultados de todos los apartados.
- Si no presentas la práctica durante la sesión, además deberás entregar la memoria de la práctica en un fichero `P8.pdf` ó `P8.doc` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.