

Laboratorio 2: Modelo lineal y mínimos cuadrados

Julián David Echeverry Correa, PhD. y Andrés Marino Álvarez Meza, PhD.

Edificio de Ingeniería Eléctrica, Oficinas 1B-136 y 1B-116

email: {jde, andres.alvarez1}@utp.edu.co

Página web del curso:

<https://sites.google.com/a/utp.edu.co/introduccion-a-la-ciencia-de-los-datos—utp-2016-2/>

1. Objetivos

- Relacionar datos utilizando modelos lineal y función de costo basada en mínimos cuadrados.
- Analizar el problema de mínimos cuadrados regularizado con base en matrices de rango deficiente o mal condicionadas.
- Extender el concepto de modelo lineal a representaciones no lineales utilizando funciones base fijas.
- Aplicar los conceptos básicos de modelado lineal utilizando mínimos cuadrados en datos sintéticos y bases de datos reales.

2. Actividades

- Consulte para que sirven los comandos `fmincon`, `fminsearch` de MatLab.
- Consulte la utilidad de la técnica TSVD (descomposición en valores singulares truncada) para la solución de problemas lineales.
- Consulte la utilidad del operador `\` de MatLab.
- Consulte las funciones `pdist` y `pdist2` de MatLab.
- Consulte los comandos `eig` y `svd` de MatLab.
- Demuestre en que consiste la solución dual del problema de mínimos cuadrados.
- Si utiliza Python, consulte las funciones equivalentes en este lenguaje a las solicitadas en las actividades para MatLab.
- Revise el código de muestra `demo_mincuadrados.m`.
- Encuentre el mínimo de $f(x)$ para la restricción $g(x) = x^3 - 8 = 0$ graficando la función de costo y la restricción en la sección 2 del código.
- Cambie la varianza del ruido blanco gaussiano en la sección 3 del código y explique que pasa con los datos.
- Cambie el dominio de los datos para la función sinc de la sección 4 del código y explique que pasa con la aproximación lineal.
- Utilice una función base polinomial y una sigmoideal para aproximar la función sinc variando los parámetros de las funciones base y la varianza del ruido.
- Realice una aproximación de alguna variable de interés de una base de datos del repositorio Kaggle. Utilice un modelo lineal con y sin regularización, y al menos dos modelos lineales con representación no lineal regularizados. Grafique las matrices de covarianza, su espectro de valores propios, y la variable aproximada para cada uno de los casos.

3. Importante

Se debe presentar un informe del laboratorio, junto con los scripts que se usaron para generar las gráficas y resultados que aparezcan en el informe. El informe debe incluir los resultados solicitados, con una pequeña discusión sobre los mismos. El informe no tiene que incluir marco teórico.