

Mínimos Cuadrados - Aplicaciones

En este proyecto se busca que el estudiante solucione algún problema de mínimos cuadrados empleando datos reales. Adicionalmente, el estudiante deberá hacer uso de la descomposición QR o la pseudo inversa vista en clase. Las aplicaciones de mínimos cuadrados pueden ser en problemas de regresión o estimación de parámetros.

Procedimiento

Asumir que se quiere modelar un conjunto de datos $\mathbf{y} \in \mathbb{R}^N$ y $\mathbf{x} \in \mathbb{R}^N$, por medio del siguiente modelo lineal en los parámetros [Boyd and Vandenberghe, 2018]

$$y = \theta_0 f_0(x) + \theta_1 f_1(x) + \cdots + \theta_P f_P(x),$$

donde las funciones $f_i(x)$ para $i = \{0, \dots, P\}$ son conocidas, por ejemplo, para un regresión polinomial $f_i(x) = x^i$. El conjunto de datos se puede modelar matricialmente así

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta},$$

donde $\mathbf{A} \in \mathbb{R}^{N \times (P+1)}$ y $\boldsymbol{\theta} \in \mathbb{R}^{P+1}$ y están definidas como

$$\mathbf{A} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_P(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_P(x_2) \\ \vdots & \vdots & & \vdots \\ f_0(x_N) & f_1(x_N) & \cdots & f_P(x_N) \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_P \end{bmatrix}.$$

Para el análisis, cuando se emplea la regresión polinomial:

- Variar $P = \{1, \dots, 5\}$.
- Para cada valor de P , tomar la medida del error medio cuadrático.
- Realizar una gráfica del error versus el numero de parámetros.
- De la figura anterior seleccionar cual es el mejor valor de P , justificar su respuesta.

Cuando x es multidimensional (Boston Housing):

- Probar para diferentes conjuntos de características.
- Tabular el error cuadrático medio para cada conjunto de parámetros.
- De la figura anterior seleccionar cual es el mejor conjunto de parámetros, justificar su respuesta.

Bases de datos

Seleccionar o construir el conjunto de datos para el problema de regresión o estimación de parámetros. Se pueden usar cualesquiera de las siguientes bases de datos:

- Boston Housing.
- Inertial Measurement Unit.
- Partículas en suspensión pm2.5.
- Otra que sea parte del proceso de investigación del estudiante.

Referencias

[Boyd and Vandenberghe, 2018] Boyd, S. and Vandenberghe, L. (2018). *Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares*. Cambridge University Press.