Section B

Monday, February 13, 2023 12:24 PM

Closed-Form Optimization

Given $J(\vec{\theta}) : \mathbb{R}^M \rightarrow \mathbb{R}$

① Write down the gradient $\nabla J(\vec{\theta})$

② Set gradient to all zeros $\nabla J(\vec{\theta}) = \begin{bmatrix} \partial J / \partial \theta_1 \\ \partial J / \partial \theta_2 \\ \vdots \\ \partial J / \partial \theta_M \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

③ Solve the system of equations for $\vec{\theta}$

④ Check whether we have a min, max, or a saddle point based on the 2nd derivatives

Closed Form Solution for Linear Regression

Notation

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} \qquad \boldsymbol{X} = \underbrace{\begin{bmatrix} x_1^{(1)} & \cdots & x_M^{(1)} \\ \vdots & & \\ \vdots & & \\ x_1^{(N)} & \cdots & x_M^{(N)} \end{bmatrix}}_{\text{Design Matrix}} \begin{array}{l} \leftarrow \vec{x}^{(1)} \\ \leftarrow \vec{x}^{(2)} \\ \vdots \\ \leftarrow \vec{x}^{(N)} \end{array}$$

① Write $J(\vec{\theta})$ in matrix/vector form

$$\begin{bmatrix} \vec{\theta}^T \vec{x}^{(1)} \\ \vec{\theta}^T \vec{x}^{(2)} \\ \vdots \\ \vec{\theta}^T \vec{x}^{(N)} \end{bmatrix} \qquad \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}$$

$$J(\vec{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left(y^{(i)} - \theta^T \vec{x}^{(i)}\right)^2$$

$$= \frac{1}{N} \frac{1}{2} \left(X\vec{\theta} - \vec{y}\right)^T \left(X\vec{\theta} - \vec{y}\right)$$

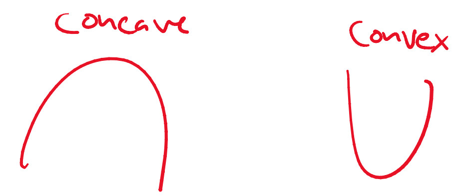
② Write gradient of $J(\vec{\theta})$

$\nabla J(\vec{\theta}) \quad X^T X \vec{\theta} \quad X^T \ldots \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

gradient of $J(\vec{\theta})$

$$\nabla J(\vec{\theta}) = \underbrace{X^T X \vec{\theta} - X^T \vec{y}}_{} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

③ Set to zero ("Normal Equations")



④ Solve for $\vec{\theta}$

$$X^T X \vec{\theta} - X^T y = 0$$

$$\Rightarrow X^T X \theta = X^T y$$

$$\Rightarrow \theta = (X^T X)^{-1} X^T y = \underset{\theta}{\operatorname{argmin}}\ J(\vec{\theta})$$

$$\boxed{\text{OH}}$$