Graded Assignment 1

Katja Bouman
10 - 10 - '17

a)

$$\vec{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \qquad X^{(i)} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$h_\theta(x) = \theta_0 x_1 + \theta_1 x_1 + \dots + \theta_n x_n$$

so the vectorized expression is:

$$h_\theta(X) = \vec{\theta}^T \cdot X^{(i)}$$

We use $\vec{\theta}^T$ because for the dot-product we need one $(m \times n)$ vector and one $(n \times m)$ vetor.

b)

The vectorized expression of the cost function is:

$$J(\vec{\theta}) = \frac{1}{2m} \sum_{i=1}^{m} (\vec{\theta}^T \cdot X^{(i)} - y^{(i)})^2$$

c)

$$\frac{\partial J(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial J(\theta)}{\partial \theta_0} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_n} \end{pmatrix} = \begin{bmatrix} \frac{1}{m} \sum_{i=1}^{m} (\vec{\theta}^T X^{(i)} - y^{(i)}) x_0^{(i)} \\ \vdots \\ \frac{1}{m} \sum_{i=1}^{m} (\vec{\theta}^T X^{(i)} - y^{(i)}) x_n^{(i)} \end{bmatrix}$$

d) The update rule of θ in the gradient descent procedure looks like this:

$$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

so the vectorized expression looks like this:

$$\theta_j := \theta_j - \frac{\alpha}{m} \sum (\vec{\theta}^T \cdot X^{(i)} - y^{(i)}) x_j^{(i)}$$