Lecture 8 (Linear Regression cotd)

1 Computing $\nabla_\theta J(\theta)$ for Linear Regression

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - \theta^T \cdot x_i)^2$$

$$\implies \nabla_\theta J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y_i - \theta^T \cdot x_i) \cdot \nabla_\theta (y_i - \theta^T \cdot x_i)$$

$$= \frac{-1}{m}\sum_{i=1}^{m} x_i(y_i - \theta^T \cdot x_i)$$

$$\therefore \theta^{(t+1)} = \theta^t + \eta\frac{1}{m}\sum_{i=1}^{m} x_i(y_i - \theta^T \cdot x_i)$$

2 Convexity

If θ_1 and θ_2 are two points in $\mathbb{D}$ (domain), then f is convex iff

$$f(\alpha\theta_1 + (1-\alpha)\theta_2) \le \alpha f(\theta_1) + (1-\alpha)f(\theta_2)$$

Strict convexity is given by strict inequality

2.1 Double Derivative of Vector (Hessian Matrix)

$$[H]_{jk} = \frac{\partial^2 f(\theta)}{\partial\theta_j \, \partial\theta_k}$$

For such a vector to be convex, H must be **positive semi-definite**. This is defined as (for a square matrix B):

$$\forall Z \in \mathbb{R}^n,\ Z^T B Z \ge 0$$