# Problem Set #1 Supervised Learning

1.

logistic regression → discriminative linear classifier

Gaussian discriminative Analysis (GDA → Generative linear Classifier

(a) $$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log\left(h_\theta(x^{(i)})\right) + (1-y^{(i)}) \log\left(1-h_\theta(x^{(i)})\right)$$

$y^{(i)} \in \{0,1\}$ $h_\theta(x) = g(\theta^T x)$

$$g(z) = \frac{1}{1+e^{-z}}$$

Show that: $z^T H z \geq 0 \quad \forall z \in \mathbb{R}^{n+1}$

↳ Hessian of $J(\theta)$

$$\frac{dg(z)}{dz} = g(z)(1-g(z))$$

$$\nabla_\theta J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \nabla_\theta \log\left(h_\theta(x^{(i)})\right) + (1-y^{(i)}) \nabla_\theta \log\left(1-h_\theta(x^{(i)})\right)$$

* $\nabla_\theta \log\left(h_\theta(x^{(i)})\right)$

$= \frac{1}{h_\theta(x^{(i)})} * h_\theta(x^{(i)}) * \left(1 - h_\theta(x^{(i)})\right) * x^{(i)}$

Hint: $\nabla_{h_\theta(x^{(i)})} \log\left(h_\theta(x^{(i)})\right) * \nabla_{\theta^T x^{(i)}} h_\theta(x^{(i)}) * \nabla_\theta \theta^T x^{(i)}$

$= \left(1 - h_\theta(x^{(i)})\right) x^{(i)}$

* $\nabla_\theta \log\left(1 - h_\theta(x^{(i)})\right)$

$= \frac{1}{1 - h_\theta(x^{(i)})} * -h_\theta(x^{(i)}) \cdot \left(1 - h_\theta(x^{(i)})\right) * x^{(i)}$

$= -h_\theta(x^{(i)}) x^{(i)}$

$\Rightarrow \nabla_\theta J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \left(1 - h_\theta(x^{(i)})\right) x^{(i)}$

$- (1 - y^{(i)}) h_\theta(x^{(i)}) x^{(i)}$

$\left(y^{(i)} - y^{(i)} h_\theta(x^{(i)}) - h_\theta(x^{(i)}) + y^{(i)} h_\theta(x^{(i)})\right) x^{(i)}$

$\left(y^{(i)} - h_\theta(x^{(i)})\right) x^{(i)}$

$$\Rightarrow \nabla_\theta J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_\theta(x^{(i)})\right)x^{(i)}$$

$$\Rightarrow \left(\nabla_\theta^2 J(\theta)\right)_k = \nabla_\theta\left(\nabla_\theta J(\theta)\right)_k$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\nabla_\theta\left(y^{(i)} - h_\theta(x^{(i)})\right)x_k^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\nabla_\theta h_\theta(x^{(i)})\,x_k^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)x^{(i)}x_k^{(i)}$$

$$\Rightarrow H = \nabla_\theta^2 J(\theta) = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)x^{(i)}x^{(i)T}$$

$$* \quad Z^T H Z = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right) Z^T x^{(i)} x^{(i)T} Z$$

$\{$is $\geq 0$ as $0 \leq h_\theta \leq 1\}$

$$Z^T x^{(i)}\left(Z^T x^{(i)}\right)^T$$

$$\left(Z^T (x^{(i)})\right)^2 \geq 0$$

$\Rightarrow$ Hence $H \geq 0$ {Positive semidefinite}

$\Rightarrow$ Hence J is Convex.