CS229: Problem Set #1

Andrew Ng

0130

Problem 1

1.
$$\frac{\partial}{\partial\theta_j}J(\theta) = \frac{\partial}{\partial\theta_j} - \frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)}\log(1 - h_\theta(x^{(i)})))$$
$$= -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\frac{1}{h_\theta(x^{(i)})} - (1 - y^{(i)})\frac{1}{1 - h_\theta(x^{(i)})}]\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)})$$

Then we calculate:

$$\frac{\partial}{\partial\theta_j}h_\theta(x^{(i)}) = \frac{\partial}{\partial\theta_j}g(\theta^T x^{(i)}) = g(\theta^T x^{(i)})(1 - g(\theta^T x^{(i)}))x_j^{(i)} = h_\theta(x^{(i)})(1 - h_\theta(x^{(i)}))x_j^{(i)}$$

And we can further simplify the above equation:

$$\frac{\partial}{\partial\theta_j}J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[(y^{(i)} - h_\theta(x^{(i)})) - (1 - y^{(i)})h_\theta(x^{(i)})]x_j^{(i)}$$
$$= -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$$

Then we can get second-order derivative:

$$H_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}J(\theta) = \frac{1}{m}\sum_{k=1}^{m} x_j\frac{\partial}{\partial\theta_i}h_\theta(x^{(k)})$$
$$= \frac{1}{m}\sum_{k=1}^{m} x_i^{(k)}x_j^{(k)}h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$

for each vector z, consider the quadratic form of Hessian matrix:

$$z^T Hz = \sum_{i=1}^{n}\sum_{j=1}^{n} z_i H_{ij} z_j = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} z_i x_i^{(k)} z_j x_j^{(k)} h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$
$$= \frac{1}{m}\sum_{k=1}^{m}(\sum_{i=1}^{n}\sum_{j=1}^{n} z_i x_i^{(k)} z_j x_j^{(k)})h_\theta(x^{(k)})(1 - h_\theta(x^{(k)}))$$
$$= \frac{1}{m}\sum_{k=1}^{m}(x^{(k)T}z)^2 h_\theta(x^{(k)})(1 - h_\theta(x^{(k)})) \geq 0 \Leftrightarrow H \succeq 0$$

2. Codes are shown in src director.

3.

1