CS229 PS1

Aryan Sood

1 Linear Classifier

1.1 (a)

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

$$\frac{\partial J}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} y^{(i)} \frac{e^{-\theta^T x^{(i)}}}{1 + e^{-\theta^T x^{(i)}}} - x_j^{(i)}(1 - y^{(i)})\frac{1}{1 + e^{-\theta^T x^{(i)}}})$$

$$\frac{\partial J}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} y^{(i)} - x_j^{(i)} h_\theta(x^{(i)})) \tag{2}$$

Now we take the derivative of (2) respect to θ_l :

$$\frac{\partial J}{\partial \theta_j \partial \theta_l} = \frac{\partial}{\partial \theta_l}(-\frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} y^{(i)} - x_j^{(i)} h_\theta(x^{(i)}))) \tag{3}$$

$$\frac{\partial J}{\partial \theta_j \partial \theta_l} = \frac{1}{m}\sum_{i=1}^{m}(x_j^{(i)} \frac{\partial}{\partial \theta_l} h_\theta(x^{(i)})) \tag{4}$$

$$\frac{\partial J}{\partial \theta_j \partial \theta_l} = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1 - h_\theta(x^{(i)}) x_j^{(i)} x_l^{(i)} \tag{5}$$

Now the Hessian is the matrix whose j-l entry is (5), now we will show that $z^T H z \geq 0$, we put $C_i = h_\theta(x^{(i)})(1 - h_\theta(x^{(i)})$ and we have $C_i \geq 0$, we see than that$z^T H z$ is equal to:

$$\sum_{j=1}^{n}\sum_{l=0}^{n}(\frac{1}{m}\sum_{i=1}^{m} C_i x_j^{(i)} x_l^{(i)}) z_j z_l \tag{6}$$

$$\sum_{i=1}^{m}\frac{1}{m} C_i \sum_{j=1}^{n}\sum_{l=1}^{n} x_j^{(i)} x_l^{(i)} z_j z_l \tag{7}$$

$$\sum_{i=1}^{m}\frac{1}{m} C_i \sum_{j=1}^{n} x_j^{(i)} z_j \sum_{l=1}^{n} x_l^{(i)} z_l \tag{8}$$

$$\sum_{i=1}^{m}\frac{1}{m} C_i (\sum_{j=1}^{n} x_j^{(i)} z_j)^2 \geq 0 \tag{9}$$

1.2 (b)

$$p(y=1|x;\phi;\mu_0;\mu_1;\Sigma) = \frac{p(x|y=1;\phi;\mu_0;\mu_1;\Sigma)p(y=1)}{p(x|y=1;\phi;\mu_0;\mu_1;\Sigma)p(y=1)+p(x|y=0;\phi;\mu_0;\mu_1;\Sigma)p(y=0)}$$

$$p(y=1|x;\phi;\mu_0;\mu_1;\Sigma) = \frac{1}{\frac{p(x|y=0;\phi;\mu_0;\mu_1;\Sigma)p(y=0)}{p(x|y=1;\phi;\mu_0;\mu_1;\Sigma)p(y=1)}+1} \tag{10}$$

Now we substitute the various formulas for the probability of x given y:

$$p(y=1|x;\phi;\mu_0;\mu_1;\Sigma) = \frac{1}{\frac{1-\phi}{\phi}exp(\frac{1}{2}(x-\mu_1)^T\Sigma(x-\mu_1)-\frac{1}{2}(x-\mu_0)^T\Sigma(x-\mu_0))+1} \tag{11}$$

We know focus on the exp:

$$(x-\mu_1)^T\Sigma(x-\mu_1)-(x-\mu_0)^T\Sigma(x-\mu_0) \tag{12}$$

$$x^T\Sigma x-\mu_1^T\Sigma x-x^T\Sigma\mu_1+\mu_1^T\Sigma\mu_1-x^T\Sigma x+\mu_0^T\Sigma x+x^T\Sigma\mu_0-\mu_0^T\Sigma\mu_0$$
$$-\mu_1^T\Sigma x-\mu_1^T\Sigma^T x+\mu_1^T\Sigma\mu_1+\mu_0^T\Sigma x+\mu_0^T\Sigma^T x-\mu_0^T\Sigma\mu_0$$
$$(\mu_1^T-\mu_0^T)(\Sigma+\Sigma^T)x+cost$$

Now we can consider $\frac{1-\phi}{phi}$ as e^a, so overall we have:

$$\frac{1}{2}(\mu_1^T-\mu_0^T)(\Sigma+\Sigma^T)x+\frac{1}{2}cost+a$$

$$\theta^T x \tag{13}$$