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 $z^T Hz \geq 0$, we put $C_i = h_\theta(x^{(i)})(1-h_\theta(x^{(i)})$ and we have $C_i \geq 0$, we see than that$z^T Hz$ 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:

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

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

2 Incomplete, Positive-Only Labels

2.1 (a)

$$P(y^i=1|x^i)=$$
$$P(y^i=1|x^i,t^{(i)}=1)\cdot P(t^{(i)}=1|x^{(i)})+P(y^i=1|x^i,t^{(i)=0})\cdot P(t^{(i)}=0|x^{(i)})=$$
$$P(y^i=1|x^i,t^{(i)}=1)\cdot P(t^{(i)}=1|x^{(i)})=$$
$$P(y^i=1|,t^{(i)}=1)\cdot P(t^{(i)}=1|x^{(i)})$$

because the positive labels are chosen uniformly we have:

$$P(y^i=1|x^i)=\alpha\cdot P(t^{(i)}=1|x^{(i)}) \tag{14}$$

point b follows directly from point a.

3 Poisson Regression

3.1 a

$$\frac{exp(ln(e^{-\lambda}\lambda^{y}))}{y!} = \frac{exp(y \cdot ln(\lambda) - \lambda)}{y!} \tag{15}$$

$$n = ln(\lambda) \tag{16}$$

$$lambda = e^{n} \tag{17}$$

$$\frac{exp(y \cdot n - e^{n})}{y!} \tag{18}$$

$$E_{poisson} = \lambda \tag{19}$$

$$E[y;x,\theta] = e^{n} = e^{\theta^{T}x} \tag{20}$$

We now find use maximum likelihood:

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{\theta^{t}x^{i}y^{i} - e^{\theta^{T}x^{i}}}}{y!}$$

$$J(\theta) = log(J(\theta)) = \sum_{i=1}^{m} -log(y!) + \sum_{i=0}^{m} (\theta^{T}x^{i}y^{i} - e^{\theta^{T}x})$$

$$\frac{\partial J}{\partial \theta_{l}} = \sum_{i=1}^{m} (x_{l}^{i}y^{i} - x_{l}^{i}e^{\theta^{T}x^{i}})$$

Stochastic gradient descent:

$$\theta_{l} := \theta_{l} + \alpha x_{l}^{i}(y^{i} - e^{\theta^{T}x^{i}})$$

4 Convexity of Generalized Linear Models

4.1 a

$$P(y;\eta) = b(y) * e^{(\eta y - a(\eta))}$$

$$\int_{-\infty}^{+\infty} b(y) * e^{(\eta y - a(\eta))} dy = 1$$

$$\frac{\partial}{\partial \eta} \int_{-\infty}^{+\infty} b(y) * e^{(\eta y - a(\eta))} dy = 0$$

$$\int_{-\infty}^{+\infty} b(y)(y - \frac{\partial}{\partial \eta} a(\eta)) e^{(\eta y - a(\eta))} dy = 0$$

$$\int_{-\infty}^{+\infty} b(y)(y) e^{(\eta y - a(\eta))} dy - \frac{\partial}{\partial \eta} a(\eta) = 0$$

$$E[y;\eta] = \frac{\partial}{\partial \eta} a(\eta)$$

4.2 b

for the variance we start with:

$$\frac{\partial}{\partial \eta} \int_{-\infty}^{+\infty} b(y) y e^{(\eta y - a(\eta))} dy = \frac{\partial}{\partial eta} (E[y;\eta]) \tag{21}$$

and we arrive at the conclusion in the same way.

4.3 c

the negative loss function is:

$$l(\eta) = -[\sum_{i=1}^{m}(b(y^i) + (\eta y^i - a(\eta)))]$$

$$l(\theta) - [\sum_{i=1}^{m}(b(y^i) + (\theta^T x^i y^i - a(\theta^T x^i)))] \frac{\partial}{\partial \theta_k} l(\theta) = -\sum_{i=1}^{m}(x_k^i y^i - \frac{\partial}{\partial \eta} a(\eta) x_k^i)$$

$$\frac{\partial^2}{\partial \theta_k \partial \theta_j} l(\theta) = \sum_{i=1}^{m}(\frac{\partial^2}{\partial \eta^2} a(\eta) x_k^i x_j^i)$$

$\frac{\partial^2}{\partial \eta^2} a(\eta)$ is always positive because is it equal to the variance. Thus by the same reasoning and manipulation of exercise 1, we obtain that the Hessian is semidefinite positive

5 Locally weighted linear regression

5.1 i

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} w^i (\theta^T x^i - y^i)^2 \tag{22}$$

$$\sum_{i=1}^{n} (\theta^T x^i - y^i) \frac{1}{2} w^i (\theta^T x^i - y^i) \tag{23}$$

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \cdot & \cdot & x_n^1 \\ x_1^2 & x_2^2 & \cdot & \cdot & x_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1^m & x_2^m & \cdot & \cdot & x_n^m \end{bmatrix} \tag{24}$$

$$2 \cdot W = \begin{bmatrix} w^1 & 0 & \cdot & \cdot & 0 \\ 0 & w^2 & \cdot & \cdot & 0 \\ 0 & 0 & w^3 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & w^m \end{bmatrix} \tag{25}$$

$$(X\theta - Y)^T W (X\theta - Y) \tag{26}$$

5.2 ii

$$\frac{\partial J(\theta)}{\partial \theta_j} = \sum_{i=1}^{m} w^i x^(i)_j (\theta^T x^i - y^i) \tag{27}$$

$$X^T 2W(X\theta - Y) = 0 \tag{28}$$

$$\theta = (X^T 2WX)^{-1} X^T Y \tag{29}$$

5.3 iii

$$\sum_{i=1}^{m} log(P(y^i; x^i, \theta)) = \sum_{i=1}^{m} log(\frac{1}{\sqrt{2\pi}\sigma^i}) - \sum_{i=1}^{m} \frac{1}{2(\sigma^i)^2}(y^i - \theta^T x^i)^2 \tag{30}$$