Name: 江才文 Student ID: 2019233157

1. Hessian in Logistic Regression (10 points)

$$\sigma'(a) = \frac{1}{(1+e^{-a})^2} e^{-a} = \frac{1}{(1+e^{-z})}\left(1 - \frac{1}{(1+e^{-z})}\right) = \sigma(a)[1-\sigma(a)] \quad ①$$

$$\frac{\partial f(w)}{\partial w} = \sum_{i=1}^{n}\left[\left(y_i \frac{1}{\sigma(w^T x_i)} - (1-y_i)\frac{1}{1-\sigma(w^T x_i)}\right)\frac{\partial}{\partial w}[\sigma(w^T x_i)]\right]$$

$$= \sum_{i=1}^{n}\left[\left(y_i \frac{1}{\sigma(w^T x_i)} - (1-y_i)\frac{1}{1-\sigma(w^T x_i)}\right)\cdot \sigma(w^T x_i)(1-\sigma(w^T x_i))\frac{\partial w^T x}{\partial w}\right.$$

$$= \sum_{i=1}^{n}\left(y_i(1-\sigma(w^T x_i)) - (1-y)\sigma(w^T x_i)\right) x_i$$

$$= \sum_{i=1}^{n}(y_i - u_i)\, x_i$$

$\cancel{\partial f}$ 继续求二阶导

$$H \frac{\partial\left(\sum_{i=1}^{n}(y_i-u_i)x_i\right)}{\partial w} = \sum_{i=1}^{n} -x_i \cdot \frac{\partial(u_i)}{\partial w}$$

$$= \sum_{i=1}^{n} -x_i\, u_i(1-u_i)\frac{\partial(w^T x)}{\partial w}$$

$$= \sum_{i=1}^{n} x_i\, u_i(1-u_i)\, x_i$$

则易知 $H = X^T S X$ ，原式得证.

2. Linear Regression (5 points)

定义线性回归模型损失函数为

$$L(w.b) = \min_{w.b} \sum (y_i - wx_i - b)^2$$

分别对 w. b 求偏导.

$$\frac{\partial L(w.b)}{\partial w} = 2\sum_{i=1}^{m} [wx_i^2 - x_i(y_i - b)]$$

$$\frac{\partial L(w.b)}{\partial b} = 2\sum_{i=1}^{m} [b - (y_i - wx_i)] = 2(mb - \sum_{i=1}^{m} (y_i - wx_i))$$

因为公式中参数 w. b. ~~m~~ m 都与 i 无关. 令 $\frac{\partial L(w.b)}{\partial w} = 0$. $\frac{\partial L(w.b)}{\partial b} = 0$

联立解得:

$$w = \frac{\sum_{i=1}^{m} y_i(x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - (\bar{x})^2} = \frac{\sum_{i=1}^{m} y_i(x_i - \frac{1}{m}\sum_{i=1}^{m} x_i)}{\sum_{i=1}^{m} x_i^2 - (\frac{1}{m}\sum_{i=1}^{m} x_i)^2}$$

$$b = \bar{y} - w\bar{x} = \frac{1}{m}\sum_{i=1}^{m} y_i - w\frac{1}{m}\sum_{i=1}^{m} x_i$$

即原式得证.

3. Gradient descent for fitting GMM (15 points)

(1)Show that the gradient of the log-likelihood wrt μ_k is

(1) $$\frac{d\,l(\theta)}{d\mu_k} = \frac{d\left(\sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)\right)}{d\mu_k}$$

$$= \sum_n \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{k'=1}^{K} \pi_{k'} N(x|\mu_{k'}, \Sigma_{k'})} \cdot \frac{d\,(\cancel{\log}\,\pi_k N(x|\mu_k, \Sigma_k))}{d\mu_k}$$

$$= \sum_n \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{k'=1}^{K} \pi_{k'} N(x|\mu_{k'}, \Sigma_{k'})} [\Sigma_k^{-1}(\cancel{x_n}\,\mu_k)]$$

$$= \sum_n r_{nk} \Sigma_k^{-1}(x_n - \mu_k)$$

(2)Derive the gradient of the log-likelihood wrt π_k without considering any constraint on π_k. (bonus: with constraint)

(2) ① without ~~Cond~~ considering $\Sigma_k \pi_k = 1$

$$\frac{d l(\theta)}{d \pi_k} = \sum_{n=1}^{N} \frac{N(x_n | u_k, \Sigma_k)}{\Sigma_j \pi_j N(x_n | u_j, \Sigma_j)}$$

② $\Sigma_k \pi = 1$ 则 需加入拉格朗日算子

$$\cancel{\frac{d l(\theta)}{d \pi_k} = \sum_{n=1}} \quad l(\theta) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) = f(\theta)$$

$$\therefore \frac{d f(\theta)}{d \pi_k} = \sum_{n=1}^{N} \frac{N(x_n | u_k, \Sigma_k)}{\Sigma_j \pi_j N(x_n | u_j, \Sigma_j)} + \lambda$$

令 $\frac{d f(\theta)}{d \pi_k} = 0$ 并两边同时乘 π_k

得 $\sum_{n=1}^{N} \frac{\pi_k N(x_n | u_k, \Sigma_k)}{\Sigma_j \pi_j N(x_n | u_j, \Sigma_j)} + \lambda \pi_k = 0$

令 $N_k + \lambda \pi_k = 0 \Longrightarrow \sum_{k=1}^{K} N_k + \lambda \sum_{k=1}^{K} \pi_k = 0$

即 $N + \lambda = 0 \qquad \lambda = -N$

$\therefore \pi_k = \frac{N_k}{N}$ 其中 $N_k = \sum_{n=1}^{N} \gamma_{nk}$

(3)Derive the gradient of the log-likelihood wrt Σ_k without considering any constraint on Σ_k. (bonus: with constraint Σ_k be a symmetric positive definite matrix.)

(3)

① $\frac{\partial l(\theta)}{\partial \Sigma_k} = \frac{\pi_k N(x_n|u_k, \Sigma_k)}{\sum_j \pi_j N(x_n|u_j, \Sigma_j)} \cdot \frac{\partial[\log(\pi_k N(x_n|u_k, \Sigma_k))]}{\partial \Sigma_j}$

$$= \sum_{n=1}^{k} \frac{\pi_k N(x_n|u_k, \Sigma_k)}{\sum_j \pi_j N(x_n|u_j, \Sigma_j)} \frac{1}{2}\left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x_n-u_k)(x_n-u_k)^T \Sigma_k^{-1}\right]$$

$$= \sum_{n=1}^{k} r_{nk} \cdot \frac{1}{2}\left[-\Sigma_k^{-1} + \Sigma_k^{-1}(x_n-u_k)(x_n-u_k)^T \Sigma_k^{-1}\right]$$

② Σ_k 是正定的

令 $\frac{\partial l(\theta)}{\partial \Sigma_k} = 0$

则可得 $\sum_{n=1}^{k} r_{nk} \Sigma_j = \sum_{n=1}^{k} r_{nk} \Sigma_j (x_n-u_k)(x_n-u_k)^T$

$$\Rightarrow \Sigma_k = \frac{\sum_{n=1}^{k} r_{nk}(x_n-u_k)(x_n-u_k)^T}{\sum_{n=1}^{k} r_{nk}}$$