DDA2020 Assignment 4

Q1.

proof: consider random variable X with n possible values $x_1, x_2, \cdots x_n$.

$$E[X] = \sum_{i=1}^{n} P(X=x_i)\cdot x_i$$

$$E[f(x)] = \sum_{i=1}^{n} P(X=x_i) f(x_i)$$

$\therefore$ we just need to show:

$$f\left(\sum_{i=1}^{n} P(X=x_i)\cdot x_i\right) \le \sum_{i=1}^{n} P(X=x_i) f(x_i)$$

use mathematical induction:

base case: $i=1$, $f(x_1) \le f(x_1)$

$i=2$, $f(P(X=x_1)x_1 + (1-P(X=x_1))x_2) \le P(X=x_1)f(x_1) + (1-P(X=x_1))f(x_2)$

the above statement is the definition of convex functions.

Inductive hypothesis:

For $n=k$: $f(P(x_1)x_1 + P(x_2)x_2 + \cdots + P(x_k)x_k) \le P(x_1)f(x_1) + \cdots + P(x_k)f(x_k)$

consider $n=k+1$: we have: $\sum_{i=1}^{k} \frac{P(x_i)}{1-P(x_{k+1})} = 1$

$$f\left(\sum_{i=1}^{k+1} P(x_i)x_i\right) = f\left((1-P(x_{k+1}))\sum_{i=1}^{k} \frac{P(x_i)}{1-P(x_{k+1})} x_i + P(x_{k+1})x_{k+1}\right)$$

$$\le (1-P(x_{k+1})) f\left(\sum_{i=1}^{k} \frac{P(x_i)}{1-P(x_{k+1})} x_i\right) + P(x_{k+1}) f(x_{k+1})$$

$$\le (1-P(x_{k+1}))\left(\sum_{i=1}^{k} \frac{P(x_i)}{1-P(x_{k+1})} f(x_i)\right) + P(x_{k+1}) f(x_{k+1}) = \sum_{i=1}^{k+1} P(x_i) f(x_i)$$

$\therefore$ Q.E.D.

Q2.

(1) Bernoulli distribution: $P(X^{(n)}|Z^{(n)}=k) = \prod_{j=1}^{d} \mu_{kj}^{x_j^{(n)}} (1-\mu_{kj})^{1-x_j^{(n)}}$

M-step:

$$\mu^{new} = \underset{\mu}{\mathrm{argmax}} \sum_{n=1}^{N} E_{q_n(z^{(n)})}\left[\ln P(z^{(n)}, x^{(n)}; \mu)\right], \ s.t. \sum_{k=1}^{K} \pi_k = 1$$

$$\mu^{new} = \underset{\mu}{\mathrm{argmax}} \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \left[\ln(P(z^{(n)}=k|\mu_{jk})) + \ln(P(x^{(n)}|z^{(n)}=k;\mu))\right]$$

$$= \underset{\mu}{\mathrm{argmax}} \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \left[\ln(\pi_k) + \sum_{j=1}^{d} \ln \mu_{kj}^{x_j^{(n)}} (1-\mu_{kj})^{1-x_j^{(n)}}\right]$$

$\therefore$ we just need:

$$\max \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \ln(\pi_k) + \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \sum_{j=1}^{d} x_j^{(n)} \ln\mu_{kj} + \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \sum_{j=1}^{d} (1-x_j^{(n)}) \ln(1-\mu_{kj})$$

$$\frac{\partial J}{\partial \mu_{kj}} = \frac{\sum_{n=1}^{N} r_k^{(n)} x_j^{(n)}}{\mu_{kj}} - \frac{\sum_{n=1}^{N} r_k^{(n)} (1-x_j^{(n)})}{1-\mu_{kj}} = 0$$

$$\Rightarrow \mu_{kj} = \frac{\sum_{n=1}^{N} r_k^{(n)} x_j^{(n)}}{\sum_{n=1}^{N} r_k^{(n)}}, \text{ the same as } \mu_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}}$$

Q.E.D.

(2) Beta distribution prior: $P(\mu_{jk}) = \frac{\mu_{jk}^{\alpha-1}(1-\mu_{jk})^{\beta-1}}{B(\alpha,\beta)}$

For MAP estimation, we try to maximum the posterior.

$$\mu^{new} = \underset{\mu}{\mathrm{argmax}} \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \left[\ln(P(z^{(n)}=k|\mu)) + \ln(P(x^{(n)}|z^{(n)}=k;\mu))\right] + \ln(P(\mu_{jk}))$$

$\therefore$ we just need:

$$\max \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \ln(\pi_k) + \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \sum_{j=1}^{d} x_j^{(n)} \ln\mu_{kj} + \sum_{n=1}^{N}\sum_{k=1}^{K} r_k^{(n)} \sum_{j=1}^{d} (1-x_j^{(n)}) \ln(1-\mu_{kj})$$
$$+ \sum_{n=1}^{N}\sum_{k=1}^{K} (\alpha-1)\ln\mu_{kj} + \sum_{n=1}^{N}\sum_{k=1}^{K} (\beta-1)\ln(1-\mu_{kj})$$

$$\therefore \frac{\partial J}{\partial \mu_{kj}} = \frac{\sum_{n=1}^{N} r_k^{(n)} x_j^{(n)} + \alpha - 1}{\mu_{kj}} - \frac{\sum_{n=1}^{N} r_k^{(n)} (1-x_j^{(n)}) + \beta - 1}{1-\mu_{kj}} = 0$$

$$\Rightarrow \mu_{kj} = \frac{\sum_{n=1}^{N} r_k^{(n)} x_j^{(n)} + \alpha - 1}{\sum_{n=1}^{N} r_k^{(n)} + \alpha + \beta - 2}, \text{ the same as } \mu_{kj} = \frac{\sum_i r_{ik} x_{ij} + \alpha - 1}{\sum_i r_{ik} + \alpha + \beta - 2}$$

$\therefore$ Q.E.D.

Q3.

proof:

$$J_W(z)=\frac{1}{2}\sum_{k=1}^{K}\sum_{i:z_i=k}\sum_{i':z_{i'}=k}(x_i-x_{i'})^2=\frac{1}{2}\sum_{k=1}^{K}\sum_{i:z_i=k}(n_kS^2+n_k(\bar{x}_k-x_i)^2)$$

$$=\frac{1}{2}\sum_{k=1}^{K}n_k^2\cdot\frac{1}{n_k}\sum_{i:z_i=k}(x_i-\bar{x}_k)^2+\frac{1}{2}\sum_{k=1}^{K}n_k\left(n_k\frac{1}{n_k}\sum_{i:z_i=k}(x_i-\bar{x}_k)^2+n_k(\bar{x}_k-\bar{x}_k)^2\right)$$

$$=\frac{1}{2}\sum_{k=1}^{K}n_k\sum_{i:z_i=k}(x_i-\bar{x}_k)^2+\frac{1}{2}\sum_{k=1}^{K}n_k\sum_{i:z_i=k}(x_i-\bar{x}_k)^2$$

$$=\sum_{k=1}^{K}n_k\sum_{i:z_i=k}(x_i-\bar{x}_k)^2$$

$\therefore$ Q.E.D.

Q4.

$$AUC=\frac{1}{m^+m^-}\sum_{i=1}^{m^+}\sum_{j=1}^{m^-}u(e_{ij})$$

we just need to prove:

$$\sum_{i=1}^{m^+}\sum_{j=1}^{m^-}u(e_{ij})=\sum_{i=1}^{m^+}rank_i-(m^+)(m^++1)/2$$

don't consider the same rank case, for ith positive sample, its rank is $rank_i$, so there are $rank_i-1$ samples have smaller predictor, including $i-1$ positive samples and $rank_i-i$ negative samples

when $1\leq j\leq rank_i-i$, $e_{ij}=1$

when $rank_i-i<j\leq n$, $e_{ij}=0$

$$\therefore \sum_{i=1}^{m^+}\sum_{j=1}^{m^-}u(e_{ij})=\sum_{i=1}^{m^+}(rank_i-i)=\sum_{i=1}^{m^+}rank_i-\frac{(1+m^+)\cdot m^+}{2}$$

if some of the data have the same g value

we take the average rank of them, since the sum of the rank don't change, so the form is still $\sum_{i=1}^{m^+}(rank_i-i)$

suppose negative samples has smaller predictors is N^-, has equal value is N_e^-. $\therefore rank_i-i=N^-+N_e^-/2$

it equals to: $\sum_{i}^{m^+}\sum_{j}^{m^-} \mathbb{I}(g(x_j^- < x_i^+)) + \frac{1}{2}\mathbb{I}(g(x_j^- = x_i^+))$

$\therefore$ Q.E.D.