AI 534 Written Assignment 0

1.a $f(z) = \log(1+z)$, $z = x^T x$, $x \in \mathbb{R}^D$

let $x = [x_1 \; x_2 \; \dots \; x_D]^T \Rightarrow x^T x = x_1^2 + x_2^2 + \dots x_D^2$

$$\nabla_x f(z) = \frac{\partial f(x)}{\partial z} \cdot \frac{\partial z}{\partial x}$$

$$\frac{\partial f(z)}{\partial z} = \frac{1}{1+z}$$

$$\frac{\partial z}{\partial x} = [2x_1, \; 2x_2, \; \dots \; 2x_D]^T$$

$$\Rightarrow \nabla_x f(z) = \frac{2}{1+z} \cdot x$$

b. $f(z) = e^{-0.5z}$ $\quad z = g(y) = y^T S^{-1} y$ $\quad y = x - \mu$ $\quad x, \mu \in \mathbb{R}^D$, $S \in \mathbb{R}^{D \times D}$

$$\nabla_x f(z) = \frac{\partial f(z)}{\partial z} \cdot \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}$$

$$\frac{\partial f}{\partial z} = -0.5 e^{-0.5z}$$

$\frac{\partial z}{\partial y}$: Let $x = [x_1 \; x_2 \; \dots \; x_D]^T \Rightarrow y = [(x_1 - \mu_1), (x_2 - \mu_2) \dots]^T$

$\mu = [\mu_1 \; \mu_2 \; \dots \; \mu_D]^T \quad = [y_1, y_2, \dots y_D]^T$

$$z = y^T S^{-1} y = [y_1, y_2 \; \dots \; y_D] \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1D} \\ s_{21} & s_{22} & \dots & s_{2D} \\ \vdots & \vdots & & \vdots \\ s_{D1} & s_{D2} & & s_{DD} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_D \end{bmatrix}$$

given that $S^{-1} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1D} \\ s_{21} & s_{22} & \dots & s_{2D} \\ \vdots & \vdots & & \vdots \\ s_{D1} & s_{D2} & \dots & s_{DD} \end{bmatrix}$

$$\Rightarrow z = \left[ (y_1 s_{11} + y_2 s_{21} + \dots y_D s_{D1}), \; (y_1 s_{12} + y_2 s_{22}, \dots y_D s_{D2}), \dots (y_1 s_{D1} + y_2 s_{D2} \dots y_D s_{DD}) \right] \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_D \end{bmatrix}$$

Date ___/___/___

$$= y_1 (y_1 S_{11} + y_2 S_{21} + \cdots y_D S_{D1}) + y_2 (y_1 S_{12} + y_2 S_{22} + \cdots y_D S_{D2}) + \cdots y_D (y_1 S_{1D} + y_2 S_{2D} + \cdots y_D S_{DD})$$

$$= \sum_{i=1}^{D} \left( y_i \left( \sum_{j=1}^{D} y_j S_{ji} \right) \right)$$

$$\frac{\partial z}{\partial y_k} = \left( \sum_{j=1}^{D} y_j S_{jk} \right) + \left( \sum_{j=1}^{D} y_j S_{kj} \right)$$

$$= \sum_{j=1}^{D} y_j ( S_{jk} + S_{kj} )$$

$$\Rightarrow \frac{\partial z}{\partial y} = \begin{bmatrix} \sum_{j=1}^{D} y_j (S_{j1} + S_{1j}) \\ \sum_{j=1}^{D} y_j (S_{j2} + S_{2j}) \\ \vdots \\ \sum_{j=1}^{D} y_D (S_{Dj} + S_{jD}) \end{bmatrix} = (S^{-1} + (S^{-1})^T) \cdot y$$

$$\frac{\partial y}{\partial x} = 1$$

$$\Rightarrow \nabla_x f(z) = -\frac{1}{2} e^{-0.5z} \cdot \begin{bmatrix} \sum_{j=1}^{D} y_j (S_{1j} + S_{j1}) \\ \vdots \\ \vdots \\ \sum_{j=1}^{D} y_j (S_{Dj} + S_{jD}) \end{bmatrix}$$

$$= -0.5 e^{-0.5z} (S^{-1} + (S^{-1})^T) \cdot y$$

Date ___/___/___

2. a. $P(\text{fair coin}) = \frac{1}{2}$.

b. $P(\text{First Toss} = H) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{10} = \frac{5}{20} + \frac{1}{20} = \frac{3}{10} = 0.3$

c. $P(\text{Fair Coin} \mid \text{Both heads}) = \dfrac{P(\text{Both Heads} \mid \text{Fair Coin}) \times P(\text{Fair Coin})}{P(\text{Both Heads})}$

$$= \frac{\left(\frac{1}{2} \times \frac{1}{2}\right) \times \frac{1}{2}}{\left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{10} \times \frac{1}{10}\right)}$$

$$= \frac{1/8}{1/8 + 1/200} = \frac{1/8}{26/200} = \frac{200}{8 \times 26} = \frac{25}{26} = 0.961$$

3. a. For the given uniform distribution $P(x;\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n} & 0 \le x \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

~~1. $\log(L(\theta)) = -n \log\theta$~~

~~$\frac{\partial \log(L(\theta))}{\partial\theta} = \frac{-n}{\theta}$ → As this a monotonically decreasing function.~~

b. $L(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$

We see that to maximize likelihood $\theta \ge \max(x_1, x_2, \ldots)$ and sufficiently small as $\frac{1}{\theta^n}$ is a decreasing function

$$\Rightarrow \theta = \max(x_1, x_2, \ldots x_n)$$

Date ___/___/___

Saathi

4. The likelihood function can be written as:

$$l = \prod_{i=1}^{N} \left( \prod_{s=a,c,g,t} \left( p_s^{I(x=s)} \right) \right)$$

$$\log(l) = \sum_{i=1}^{N} \left( \sum_{s=a,c,g,t} I(x=s) \cdot \log p_s \right)$$

Writing the Lagrangian equation

$$L = \sum_{i=1}^{N} \left( \sum_{s=a,c,g,t} I(x=s) \cdot \log p_s \right) - \lambda \left( p_a + p_c + p_g + p_t - 1 \right)$$

$$\frac{\partial L}{\partial p_s}_{\forall s \in \{a,c,g,t\}} = \sum_{i=1}^{N} \left( \frac{1}{p_s} \cdot I(x=s) \right) - \lambda \cdot 1 = 0$$

as no. of occurances of $A = n_a$

$G = n_g$

$C = n_c$

$T = n_t$

$$\Rightarrow \frac{\partial L}{\partial p_s} = \left( \frac{n_s}{p_s} - \lambda \right) \text{ for } s \in \{a,c,g,t\}.$$

$$= 0$$

$$\Rightarrow \lambda = \frac{n_s}{p_s} \Rightarrow \frac{n_a}{p_a} = \frac{n_c}{p_c} = \frac{n_g}{p_g} = \frac{n_t}{p_t} = \lambda$$

$$\Rightarrow n_a = \lambda p_a, \quad n_c = \lambda p_c, \quad n_g = \lambda p_g, \quad n_t = \lambda p_t$$

$$n_a + n_g + n_c + n_t = N = \lambda (p_a + p_c + p_g + p_t) = N$$

$$\Rightarrow \lambda = \frac{N}{p_a + p_c + p_g + p_t}$$

But $p_a + p_c + p_g + p_t = 1$

$$\Rightarrow \frac{n_a}{p_a} = \frac{n_c}{p_c} = \frac{n_g}{p_g} = \frac{n_t}{p_t} = N$$

$$\Rightarrow p_a = \frac{n_a}{N}, \quad p_c = \frac{n_c}{N}, \quad p_g = \frac{n_g}{N}, \quad p_t = \frac{n_t}{N}$$

Date ___/___/______

5.a. Expected Loss $= 0.4 \times 0. + 0.6 \times 5 = 3.$

b. Lets assume that $P(\hat{y}=1) = p$

$\Rightarrow$ Expected Loss $= 0.4(p.0 + (1-p)5) + 0.6(p.10 + (1-p)0)$

$= 2 - 2p + 6p = 4p + 2$

We minimize the expected loss when $p = 0$.

$\Rightarrow$ The best solution is to always predict 0

$\Rightarrow p = P(\hat{y}=1) = 0.$

c. Again let $P(\hat{y}=1) = p$.

~~$\Rightarrow$ Expected Loss = 4p+2~~

Expected Loss $= P(y=1)(p.0 + (1-p).5) + P(y=0) \times (p.10 + (1-p).0)$

$= 5P(y=1) - 5pP(y=1) + 10pP(y=0)$

$= 5P(y=1) - 5pP(y=1) + 10p - 10pP(y=1)$

$= 10p - 15pP(y=1) + 5P(y=1)$

$\frac{d.E.L}{dp} = 10 - 15P(y=1)$ (EL = Exp. Loss)

We can observe that $\frac{dE.L}{dp}$ is decreasing as $P(y=1)$ increases.

At $P(y=1) = \frac{2}{3}$ ; $\frac{dEL}{dp} = 0$

~~$\Rightarrow$ The expected loss is a concave function, that achieves a maximum when $P(y=1) = \frac{2}{3}$~~

~~$\Rightarrow$ E.L is concave with respect to p~~

$\Rightarrow$ When $P(y=1) = 0$ , $\frac{dEL}{dp} = 10$

When $P(y=1) = 1$ , $\frac{dEL}{dp} = -5$

When $P(y=1) = \frac{2}{3}$ , $\frac{dEL}{dp} = 0$

Thus when $P(y=1) < \frac{2}{3}$, the expected loss increases with $p$

Thus to minimize loss, $p = 0$

When $P(y=1) \geq \frac{2}{3}$, the expected loss decreases with increase of $p$.

Thus to minimize loss $p = 1$

At $P(y=1) = \frac{2}{3}$, $p$ does not affect the loss

$\Rightarrow$ If $P(y=1) \geq \frac{2}{3}$ optimally $p = 1 = P(\hat{y}=1)$

If $P(y=1) < \frac{2}{3}$ optimally $p = 0$

Date __/__/__

d Consider the loss matrix :

| $\hat{y}$ \ $y$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | a | 0 |

$\Rightarrow$ Expected Loss $= P(y=1)\left((1-p)\cdot a\right) + P(y=0)\cdot p$

where $p = P(\hat{y}=1)$

$$\Rightarrow E.L = aP(y=1) - apP(y=1) + pP(y=0)$$
$$= aP(y=1) - apP(y=1) + p - pP(y=1)$$
$$= p + aP(y=1) - apP(y=1) - pP(y=1)$$

To be the threshold $\frac{dE.L}{dp} = 0$

$$\Rightarrow 1 - aP(y=1) - P(y=1) = 0$$
$$\Rightarrow 1 - P(y=1)\times(a+1) = 0$$
$$\Rightarrow P(y=1)\times(a+1) = 1$$

But we know $P(y=1) = 0.1$ at the threshold

$$\Rightarrow 0.1\times(a+1) = 1$$
$$a+1 = 10$$
$$\Rightarrow \underline{\underline{a = 9}}$$

The matrix is :

| $\hat{y}$ \ $y$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 9 | 0 |