Chang Chen (cc992)
Lingyu Zhou (lz568)

CS 5780 Homework 3

Problem 1

1(i) $P(Y=1) = \boxed{\frac{3}{8}}$ $\quad P(Y=0) = 1 - P(Y=0) = 1 - \frac{3}{8} = \boxed{\frac{5}{8}}$

$$\mathbb{E}[C \mid Y=0] = \frac{(1+1+0+1+0)}{5} = \boxed{\frac{3}{5}}$$

$$Var(C \mid Y=0) = \frac{(1-\frac{3}{5})^2 + (1-\frac{3}{5})^2 + (0-\frac{3}{5})^2 + (1-\frac{3}{5})^2 + (0-\frac{3}{5})^2}{5} = \boxed{\frac{6}{25}}$$

$$\mathbb{E}[C \mid Y=1] = \frac{1+0+0}{3} = \boxed{\frac{1}{3}}$$

$$Var(C \mid Y=1) = \frac{(1-\frac{1}{3})^2 + (0-\frac{1}{3})^2 + (0-\frac{1}{3})^2}{3} = \boxed{\frac{2}{9}}$$

$$\mathbb{E}[D \mid Y=0] = \frac{3+2+4+3+5}{5} = \boxed{\frac{17}{5}}$$

$$Var(D \mid Y=0) = \frac{(3-\frac{17}{5})^2 + (2-\frac{17}{5})^2 + (4-\frac{17}{5})^2 + (3-\frac{17}{5})^2 + (5-\frac{17}{5})^2}{5} = \boxed{\frac{26}{25}}$$

$$\mathbb{E}[D \mid Y=1] = \frac{5+7+6}{3} = \boxed{6}$$

$$Var(D \mid Y=1) = \frac{(5-6)^2 + (7-6)^2 + (6-6)^2}{3} = \boxed{\frac{2}{3}}$$

1(ii)

$$P(Y=0 \mid C=0, D=2) = \frac{P(C=0, D=2 \mid Y=0) * P(Y=0)}{P(C=0, D=2)}$$

$$P(C=0, D=2 \mid Y=0) = P(C=0 \mid Y=0) * P(D=2 \mid Y=0)$$

$$= \frac{1}{\sqrt{\frac{6}{25}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-\frac{3}{5}}{\sqrt{\frac{6}{25}}}\right)^2} * \frac{1}{\sqrt{\frac{26}{25}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{2-\frac{17}{5}}{\sqrt{\frac{26}{25}}}\right)^2}$$

$$= 0.0889598379$$

$$P(C=0, D=2) = P(C=0, D=0 \mid Y=1) \cdot P(Y=1) + P(C=0, D=0 \mid Y=0) \cdot P(Y=0)$$

$$= \underbrace{P(C=0 \mid Y=1) \cdot P(D=2 \mid Y=1)}_{\frac{1}{\sqrt{\frac{2}{9}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-\frac{1}{3}}{\sqrt{\frac{2}{9}}}\right)^2} * \frac{1}{\sqrt{\frac{2}{3}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{2-6}{\sqrt{\frac{2}{3}}}\right)^2}} \cdot \underbrace{P(Y=1)}_{\frac{3}{8}} + \underbrace{P(C=0 \mid Y=0) \cdot P(D=2 \mid Y=0)}_{\substack{0.0889598379 \\ \text{(calculated above)}}} \cdot \underbrace{P(Y=0)}_{\frac{5}{8}}$$

$$\frac{1}{\sqrt{\frac{2}{9}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-\frac{1}{3}}{\sqrt{\frac{2}{9}}}\right)^2} * \frac{1}{\sqrt{\frac{2}{3}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{2-6}{\sqrt{\frac{2}{3}}}\right)^2} \Big\} \; 1.97863011 * 10^{-6}$$

$$= 7.41986291 * 10^{-7} + 0.0555989869$$

$$\approx 0.05560064068$$

$$P(Y=0 \mid C=0, D=2) = \frac{0.0889598379 * \frac{5}{8}}{0.05560064068} \approx \boxed{0.999986655}$$

$$P(Y=1 \mid C=0, D=2) = \frac{P(C=0, D=2 \mid Y=1) * P(Y=1)}{P(C=0, D=2)}$$

$$P(C=0, D=2 \mid Y=1) = P(C=0 \mid Y=1) * P(D=2 \mid Y=1)$$

$$= \frac{1}{\sqrt{\frac{2}{9}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-\frac{1}{3}}{\sqrt{\frac{2}{9}}}\right)^2} * \frac{1}{\sqrt{\frac{2}{3}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{2-6}{\sqrt{\frac{2}{3}}}\right)^2}$$

$$= 1.97863011 * 10^{-6}$$

$$P(Y=1 \mid C=0, D=2) = \frac{1.97863011 * 10^{-6} * \frac{3}{8}}{0.05560064068} \simeq \boxed{1.33449234 * 10^{-5}}$$

This Gaussian Naive Bayes model would predict that bike lane is obstructed for a Tuesday where there is no construction.

1 (iii)

$$P(C=0|Y=0) = \frac{2+1}{5+2} = \boxed{\frac{3}{7}}$$

$$P(C=1|Y=0) = \frac{3+1}{5+2} = \boxed{\frac{4}{7}}$$

$$P(C=0|Y=1) = \frac{2+1}{3+2} = \boxed{\frac{3}{5}}$$

$$P(C=1|Y=1) = \frac{1+1}{3+2} = \boxed{\frac{2}{5}}$$

$$P(D=0|Y=0) = \frac{5+1}{5+2} = \boxed{\frac{6}{7}}$$

$$P(D=1|Y=0) = \frac{0+1}{5+2} = \boxed{\frac{1}{7}}$$

$$P(D=0|Y=1) = \frac{2+1}{3+2} = \boxed{\frac{3}{5}}$$

$$P(D=1|Y=1) = \frac{1+1}{3+2} = \boxed{\frac{2}{5}}$$

1 (iv)

Seyab is not correct, because the Bayes Optimal classifier minimizes the the expected misclassification cost (error rate). Thus, it might not necessarily minimize the expected time.

P.2

$$\mathbb{P}(y|x_i) = \left(1+e^{-[y(w^T x_i)]}\right)^{-1} \quad \text{--- (1)}$$

(i). $\mathbb{P}(y_i=1|x_i) + \mathbb{P}(y_i=-1|x_i) = 1$

plug in (1), we have: $\mathbb{P}(y_i=1|x_i) = (1+e^{-w^T x_i})^{-1}$

$\mathbb{P}(y_i=-1|x_i) = (1+e^{w^T x_i})^{-1}$

Let $w^T x_i = s$, then:

$$\mathbb{P}(y_i=1|x_i) + \mathbb{P}(y_i=-1|x_i) = \frac{1}{1+e^{-s}} + \frac{1}{1+e^{s}}$$

$$= \frac{1}{1+\frac{1}{e^s}} + \frac{1}{1+e^s}$$

$$= \frac{1}{\frac{e^s+1}{e^s}} + \frac{1}{1+e^s}$$

$$= \frac{e^s}{e^s+1} + \frac{1}{1+e^s}$$

$$= \frac{1+e^s}{1+e^s} = 1 \qquad \boxtimes$$

(ii).

$$LHS = \sigma'(s) = d[(1+e^{-s})^{-1}]/dx = -(1+e^{-s})^{-2}\cdot(-e^{-s}) = \frac{e^{-s}}{(1+e^{-s})^2}$$

$$RHS = \sigma(s)[1-\sigma(s)] = \frac{1}{1+e^{-s}}\left(\frac{e^{-s}}{1+e^{-s}}\right) = \frac{e^{-s}}{(1+e^{-s})^2} = LHS \qquad \boxtimes.$$

(iii).

$$\nabla_w \ln[\mathbb{P}(\vec{y}|X,\vec{w})] = \frac{\partial \ln[\mathbb{P}(\vec{y}|X,\vec{w})]}{\partial w} = \frac{\partial \ln[\sigma(\vec{y}(\vec{w}^T X))]}{\partial w}, \text{ where } s = \vec{y}(\vec{w}^T X)$$

$$= \frac{1}{\sigma(\vec{y}(\vec{w}^T X))} \cdot \frac{\partial \sigma(\vec{y}(\vec{w}^T X))}{\partial w} \text{ (chain rule)}$$

$$= \frac{1}{\sigma(\vec{y}(\vec{w}^T X))} \cdot \frac{\partial \sigma(\vec{y}(\vec{w}^T X))}{\partial(\vec{y}(\vec{w}^T X))} \cdot \frac{\partial(\vec{y}(\vec{w}^T X))}{\partial \vec{w}} \quad \left(\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z}\cdot\frac{\partial z}{\partial x}\right)$$

$$= \frac{1}{\cancel{\sigma(\vec{y}(\vec{w}^T X))}} \cdot \sigma\cancel{(\vec{y}(\vec{w}^T X))}[1-\sigma(\vec{y}(\vec{w}^T X))]\, X\cdot\vec{y}$$

$$= [1-\sigma(\vec{y}(\vec{w}^T X))] \cdot X \cdot \vec{y}$$

$$= \sum_{i=1}^{n} [1-\sigma(y_i(w^T X_i))] \cdot y_i X_i \quad \text{(unpack matrix multiplication)} \quad \boxtimes.$$

(iv).

Base case:

$W_0 = \vec{0} = \sum_{i=1}^{n} 0 \, x_i$, $0 \in \mathbb{R}$, since proved

General case:

Assume $W_k = \sum_{i=1}^{n} c_i x_i$ is true, then we have,

$$W_{k+1} = W_k + (-\eta \nabla W_k)$$

$$= W_k - \eta \nabla W_k$$

$$= \sum_{i=1}^{n} c_i \vec{x_i} - \eta \sum_{i=1}^{n} (1-\sigma(y_i(\vec{w_k}^T \vec{x_i}))) \, y_i \vec{x_i}$$

We know $\sigma(\cdot)$ takes in a scalar and output a scalar, that is $\sigma(\cdot) \in \mathbb{R}$ and since $y_i = -1$ or $+1$, is also a scalar. we know $(1-\sigma(y_i(\vec{w_k}^T \vec{x_i}))) y_i \in \mathbb{R}$. Note that is scalar might be related to $\vec{x_i}$ BUT it's definitely a scalar, we can then make $(1-\sigma(y_i(\vec{w_k}^T \vec{x_i}))) = d_i$ $(d_i \sim \vec{x_i})$. There fore we have

$$W_{k+1} = \cdots = \sum_{i=1}^{n} c_i \vec{x_i} - \eta \sum_{i=1}^{n} d_i \vec{x_i}$$

$$= \sum_{i=1}^{n} (c_i - \eta d_i) \vec{x_i}$$

$$= \sum_{i=1}^{n} \tilde{c_i} \vec{x_i} \quad (\tilde{c_i} = c_i - \eta d_i)$$

Hence proof is done by completion of both base case and general case $\boxtimes$

Problem 3

3 (i) $f(w) = 5(w-11)^4 \qquad \alpha = \frac{1}{40} \qquad w_0 = 13$

Step 1: $S = -\alpha \nabla f(w) \qquad \nabla f(w) = 20(w-11)^3$

$$= -\frac{1}{40}\left(20(w-11)^3\right)$$

$$= -\frac{1}{40}\left(20(13-11)^3\right)$$

$$= -4$$

$$w_1 = w_0 + S = 13 - 4 = \boxed{9}$$

Step 2: $S = -\frac{1}{40}\left(20(9-11)^3\right)$

$$= 4$$

$$w_2 = w_1 + S = 9 + 4 = \boxed{13}$$

3 (ii) Adagrad Algorithm

$w_d^0 = 13 \qquad Z_d = 0 \qquad g = \nabla f(w) = 20(w-11)^3$

$$Z_d \leftarrow Z_d + g_d^2$$

$$w_d^{t+1} \leftarrow w_d^t - \alpha \frac{g_d}{\sqrt{Z_d + \varepsilon}}$$

Iteration 1:

$$Z_d^1 \leftarrow 0 + \left(20(13-11)^3\right)^2 = 25600$$

$$w_d^1 \leftarrow 13 - \frac{1}{40}\left(\frac{160}{\sqrt{25600+0}}\right) \simeq \boxed{12.975}$$

Iteration 2:

$$z_d^2 \leftarrow 25600 + \left(20(12.975 - 11)^3\right)^2 = 49339.00933$$

$$w_d^1 \leftarrow 12.975 - \frac{1}{40}\left(\frac{154.0746878}{\sqrt{49339.00933 + 0}}\right) \simeq \boxed{12.95766892}$$

3(iii) $\nabla f(w) = 20(w-11)^3$

$\alpha = \frac{1}{20}$

I1: $S = -\frac{20}{20}(13-11)^3 = -8$

$w_1 = 13 - 8 = 5$

I2: $S = -\frac{20}{20}(5-11)^3 = 216$

$w_2 = 5 + 216 = 221$

I3: $S = -\frac{20}{20}(221-11)^3 = -9261000$

$w_3 = 221 - 9261000 = -9260779$

$\alpha = \frac{1}{160}$

$S = -\frac{20}{160}(13-11)^3 = -1$

$w_1 = 13 - 1 = 12$

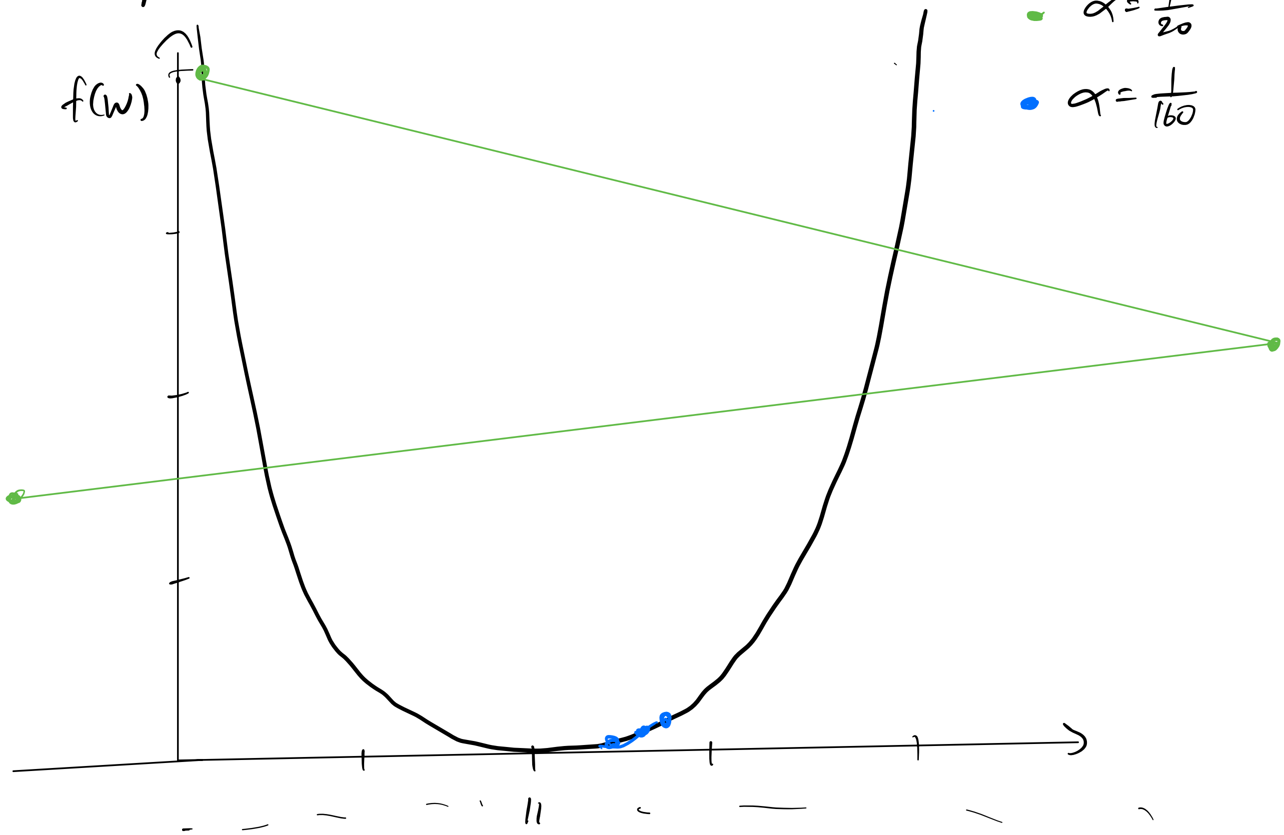
$S = -\frac{20}{160}(12-11)^3 = -0.125$

$w_2 = 12 - 0.125 = 11.875$

$S = -\frac{20}{160}(11.875 - 11)^3 = -0.0837402344$

$w_3 = 11.875 - 0.0837402344 = 11.79125977$

Graph



The graph shows that $\alpha = \frac{1}{20}$ is converging at a faster rate than $\alpha = \frac{1}{160}$.