CSE584 HOMEWORK 2

CHEN SUN
CXS1031@PSU.EDU

1. Question 1

```
# python 2.7
def hw1(u, lambdavar):
        # compute u_square and lambda_square
        # set both to be float value for division
        u_square = float(u*u)
        lambda_square = float(lambdavar*lambdavar)

        #Case 1: when x = 0, the value of function:
        f1 = u_square
        #initialize
        argmin_x = 0
        min_f = f1

        #Case 2: when x > 0, the argmin_x is:
        x2 = float(u) - float(lambdavar)/2
        f2 = u * lambdavar - lambda_square/4 #lambda_square is float value

        if x2 > 0: # if x1 <= 0, Case 2 can not be considered
                if f2 < min_f:
                        argmin_x = x2
                        min_f = f2

        #Case 3: when x < 0, the argmin_x is
        x3 = float(u) + float(lambdavar)/2
        f3 = -1 * u * lambdavar - lambda_square/4 #lambda_square float value

        if x3 < 0:
                if f3 < min_f:
                        argmin_x = x3
                        min_f = f3

        return argmin_x
```

2. Question 2

Since X, Y and C are all binary variables, we have the following:

(1)

$$
\begin{aligned}
P(C|X,Y) &= \frac{P(C,X,Y)}{P(X,Y)} \\
&= \frac{P(Y|X)P(X|C)P(C)}{P(Y|X)P(X|C)P(C) + P(Y|X)P(X|\neg C)P(\neg C)} \\
&= \frac{P(X|C)P(C)}{P(X|C)P(C) + P(X|\neg C)P(\neg C)}
\end{aligned}
$$

(2)

$$
\begin{aligned}
P(C|X) &= \frac{P(X|C)P(C)}{P(X)} \\
&= \frac{P(X|C)P(C)}{P(X|C)P(C) + P(X|\neg C)P(\neg C)}
\end{aligned}
$$

(3)

$$
\begin{aligned}
P(Y) = {} & P(Y|X)P(X|C)P(C) + P(Y|X)P(X|\neg C)P(\neg C) \\
& + P(Y|\neg X)P(\neg X|C)P(C) + P(Y|\neg X)P(\neg X|\neg C)P(\neg C)
\end{aligned}
$$

$$
\begin{aligned}
P(Y,C) &= P(X,Y,C) + P(\neg X,Y,C) \\
&= P(Y|X)P(X|C)P(C) + P(Y|\neg X)P(\neg X|C)P(C) \\
P(C|Y) &= \frac{P(Y,C)}{P(Y)}
\end{aligned}
$$

3. Question3

(1) $\frac{2}{3}$

(2) $\frac{1}{3}$

(3) $\frac{50}{77}$

(4) $\frac{9}{29}$

(5)

$$
\begin{aligned}
P(Y=1) &= P(Y|X)P(X|C)P(C) + P(Y|X)P(X|\neg C)P(\neg C) \\
&\quad + P(Y|\neg X)P(\neg X|C)P(C) + P(Y|\neg X)P(\neg X|\neg C)P(\neg C) \\
\\
&= P(Y=1|X=1)P(X=1|C=0)P(C=0) + P(Y=1|X=1)P(X=1|C=1)P(C=1) \\
&\quad + P(Y=1|X=0)P(X=0|C=1)P(C=1) + P(Y|X=0)P(X=0|C=0)P(C=0) \\
\\
&= \frac{3}{5}\cdot\frac{2}{5}\cdot\frac{1}{3} + \frac{3}{5}\cdot\frac{4}{9}\cdot\frac{2}{3} + \frac{2}{3}\cdot\frac{5}{9}\cdot\frac{2}{3} + \frac{2}{3}\cdot\frac{3}{5}\cdot\frac{1}{3} \\
\\
&= \frac{1292}{2025}
\end{aligned}
$$

$$
\begin{aligned}
P(Y=1, C=0) &= P(X,Y,C) + P(\neg X, Y, C) \\
&= P(Y|X)P(X|C)P(C) + P(Y|\neg X)P(\neg X|C)P(C) \\
&= P(Y=1|X=1)P(X=1|C=0)P(C=0) + P(Y=1|X=0)P(X=0|C=0)P(C=0) \\
&= \frac{3}{5}\cdot\frac{2}{5}\cdot\frac{1}{3} + \frac{2}{3}\cdot\frac{3}{5}\cdot\frac{1}{3} \\
&= \frac{16}{75}
\end{aligned}
$$

$$
\begin{aligned}
P(C=0|Y=1) &= \frac{P(Y=1, C=0)}{P(Y=1)} \\
&= \frac{108}{323}
\end{aligned}
$$

4. Question4

(1) Since x_i are independent of each other:

$$
\begin{aligned}
\log P(D|p) &= \log(P(x_1|p)P(x_2|p)\cdots P(x_n|p)) \\
&= \log[p^{x_1-4}(1-p)\cdot p^{x_2-4}(1-p)\cdots p^{x_n-4}(1-p)] \\
&= \log[p^{(\sum\limits_{i=1}^{n} x_i)-4n}(1-p)^n] \\
&= [(\sum_{i=1}^{n} x_i) - 4n]\log p + n\log(1-p)
\end{aligned}
$$

(2) Denote the base of logorithm as a.

$$
\begin{aligned}
\frac{\partial \log P(D|p)}{\partial p} &= \frac{(\sum\limits_{i=1}^{n} x_i) - 4n}{p\ln a} - \frac{n}{(1-p)\ln a} \\
&= \frac{[(\sum\limits_{i=1}^{n} x_i) - 4n](1-p) - np}{p(1-p)\ln a} \\
&= \frac{(\sum\limits_{i=1}^{n} x_i) - 4n - (\sum\limits_{i=1}^{n} x_i p) + 3np}{p(1-p)\ln a}
\end{aligned}
$$

For all $0 < p < 1$, i.e. $p(1-p) \neq 0$, to find the optimal parameter p:

Let $\dfrac{\partial \log P(D|p)}{\partial p} = 0$.

We have $p = \dfrac{(\sum_{i=1}^{n} x_i) - 4n}{(\sum_{i=1}^{n} x_i) - 3n}$

5. Question5

For any vector $\vec{x} = \begin{pmatrix} a \\ b \end{pmatrix}$

$$(a,b)\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = a^2 - ab + b^2 > 0 \qquad \text{for all non-zero } \vec{x} \tag{1}$$

So it is positive definite and positive semidefinite.

$$(a,b)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = 2ab \tag{2}$$

Since we can not decide the sign, it is neither positive semidefinite nor positive definite.

$$(a,b)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = (a+b)^2 \tag{3}$$

if $a = -b$, the value is 0; otherwise the value is greater than 0.
So it is positive semidefinite.

6. Question 6

(1) Suppose $\vec{x_1}, \vec{x_2} \in A$. Then for any $0 \le \theta \le 1$, we have :

$$\begin{aligned} B(\theta\vec{x_1} + (1-\theta)\vec{x_2}) &= \theta B\vec{x_1} + (1-\theta)B\vec{x_2} \\ &= \theta c + (1-\theta)c \\ &= c \end{aligned}$$

So $\theta\vec{x_1} + (1-\theta)\vec{x_2}$ is also in A. A is a convex set.

(2) Suppose $\vec{x_1}, \vec{x_2} \in A$. Then for any $0 \le \theta \le 1$, we have :

$$\begin{aligned} B(\theta\vec{x_1} + (1-\theta)\vec{x_2}) &= \theta B\vec{x_1} + (1-\theta)B\vec{x_2} \\ &\le \theta c + (1-\theta)c \\ &= c \end{aligned}$$

So $\theta\vec{x_1} + (1-\theta)\vec{x_2}$ is also in A. A is a convex set.

(3) Suppose $\vec{x_1}, \vec{x_2} \in A$. Then for any $0 \le \theta \le 1$, we have :

$$\begin{aligned} g(\theta\vec{x_1} + (1-\theta)\vec{x_2}) &\le \theta g(\vec{x_1}) + (1-\theta)g(\vec{x_2}) \\ &\le \theta c + (1-\theta)c \\ &= c \end{aligned}$$

So $\theta\vec{x_1} + (1-\theta)\vec{x_2}$ is also in A. A is a convex set.

(4) Since x^2 is a convex function.
Suppose $\vec{x_1}, \vec{x_2} \in A$. Then for any $0 \le \theta \le 1$, we have :

$$\begin{aligned}(\theta\vec{x_1} + (1-\theta)\vec{x_2})^2 &\le \theta(\vec{x_1})^2 + (1-\theta)(\vec{x_2})^2 \\ &\le \theta c + (1-\theta)c \\ &= c\end{aligned}$$

So $\theta\vec{x_1} + (1-\theta)\vec{x_2}$ is not always in A. So A is not a convex set.

7. Question 7

(1) For any two vectors $\vec{x} = (x_1, x_2, \cdots, x_k)$ and $\vec{y} = y_1, y_2, \cdots, y_n$ in the domain of f:

$$\begin{aligned}f(\theta\vec{x} + (1-\theta)\vec{y}) &= \sum_{i=1}^{n} |\theta\vec{x_i} + (1-\theta)\vec{y_i}| \\ &\le \sum_{i=1}^{n}(|\theta\vec{x_i}|) + \sum_{i=1}^{n}(|(1-\theta)y_i|) \\ &= \theta\sum_{i=1}^{n}|\vec{x}| + (1-\theta)\sum_{i=1}^{n}|\vec{y}| \\ &= \theta f(\vec{x}) + (1-\theta)f(\vec{y})\end{aligned}$$

So $f(\vec{x})$ is a convex function.

(2)

$$\frac{\partial^2 f}{\partial x^2} = \frac{e^{-x}(1-e^{-x})}{(1+e^{-x})^3}\begin{cases} > 0 & \text{if } x < 0 \\ \le 0 & \text{otherwise}\end{cases}$$

So $f(x)$ is not convex, and $f(x)$ is convex only when $x > 0$.

(3) Suppose the base of logarithm is a.

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{(1+e^{-x})\cdot(-e^{-x})}{\ln a} \\ \frac{\partial^2 f}{\partial x^2} &= e^{-x} + 2e^{-2x} \ge 0\end{aligned}$$

So $f(x)$ is convex.

(4) Suppose the base of logarithm is a.

$$\frac{\partial f}{\partial \vec{x}} = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}]^T$$
$$= [-\frac{e^{-2x_1}}{\ln a}, -\frac{e^{-2x_2}}{\ln a}]^T$$

$$\frac{\partial^2 f}{\partial \vec{x}^2} = \begin{pmatrix} 2e^{-2x_1} & 0 \\ 0 & 2e^{-2x_2} \end{pmatrix}$$

For an arbitary vector $\vec{v} = (a, b)^T$

$$\vec{v}^T \frac{\partial^2 f}{\partial \vec{x}^2} \vec{v} = 2e^{-2x_1} a^2 + 2e^{-2x_2} b^2 \geq 0$$

So $\dfrac{\partial^2 f}{\partial \vec{x}^2}$ is positive semidefinite.

So $f(\vec{x})$ is convex.

8. Question 8

Suppose $f = \vec{x}^T A \vec{x}$, we want to prove $\dfrac{\partial f}{\partial \vec{x}} = (A + A^T)\vec{x}$.

Since $\vec{x} = (x_1, x_2, ..., x_n)$, $f = \vec{x}^T A \vec{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$

$\dfrac{\partial f}{\partial x_k} = \sum_{i-1}^{n} a_{ki} x_i + \sum_{j=1}^{n} a_{jk} x_j$ for all k = 1,2...n

so $\dfrac{df}{d\vec{x}} = A^T \vec{x} + A\vec{x} = (A^T + A)\vec{x}$.