Homework 1

VE471 - Introduction to Data Science Spring 2022

* Name: Huang Yucheng ID: 519021910885

1.1

1.1.1

$$B = \begin{pmatrix} 2 & 2 & 2 \\ 5 & 8 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 10 \end{pmatrix}$$

1.1.2

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 10 \end{pmatrix}$$

1.1.3

$$AB = \begin{pmatrix} 9 & 12 & 4 \\ 7 & 12 & 15 \\ 0 & 0 & 130 \end{pmatrix}$$

We need to calculate $(AB)^{-1}$

$\begin{vmatrix} 9 & 12 & 4 \\ 7 & 12 & 15 \\ 0 & 0 & 100 \end{vmatrix} = (0)(-1)^{3+1}\begin{vmatrix} 12 & 4 \\ 12 & 15 \end{vmatrix} + (0)(-1)^{3+2}\begin{vmatrix} 9 & 4 \\ 7 & 15 \end{vmatrix} + (100)(-1)^{3+3}\begin{vmatrix} 9 & 12 \\ 7 & 12 \end{vmatrix}$

$= 100\begin{vmatrix} 9 & 12 \\ 7 & 12 \end{vmatrix} = 2400$

The adjugate matrix is $\begin{pmatrix} 1200 & -1200 & 132 \\ -700 & 900 & -107 \\ 0 & 0 & 24 \end{pmatrix}$

The inverse matrix is the adjugate matrix divided by the determinant.

Thus, the inverse matrix is $\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{11}{200} \\ -\frac{7}{24} & \frac{3}{8} & -\frac{107}{2400} \\ 0 & 0 & \frac{1}{100} \end{pmatrix}$

$\overrightarrow{v_2} = (AB)^1 \cdot \vec{x} = \begin{pmatrix} \frac{11}{2} \\ \frac{53}{24} \\ 1 \end{pmatrix}$

1.2

1.2.1

$$\sigma(-x) = \frac{1}{1+e^x}$$

$$\begin{aligned} 1 - \sigma(x) &= 1 - \frac{1}{1+e^{-x}} \\ &= 1 - \frac{e^x}{e^x+1} \\ &= \frac{1}{e^x+1} \end{aligned}$$

1.2.2

$$\begin{aligned}
\frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) &= -\frac{1}{(1+e^{-x})^2}\frac{d}{dx}\left(1+e^{-x}\right) \\
&= -\frac{1}{(1+e^{-x})^2}\left(-e^{-x}\right) \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{e^{x}}{(1+e^{x})^2}
\end{aligned}$$

$$\begin{aligned}
\sigma(x)(1-\sigma(x)) &= \sigma(x)\cdot\sigma(-x) \\
&= \frac{e^x}{1+e^x}\cdot\frac{1}{1+e^x} \\
&= \frac{e^x}{(1+e^x)^2}
\end{aligned}$$

$$\frac{d\sigma(x)}{dx} = \sigma(x)(1-\sigma(x))$$

1.2.3

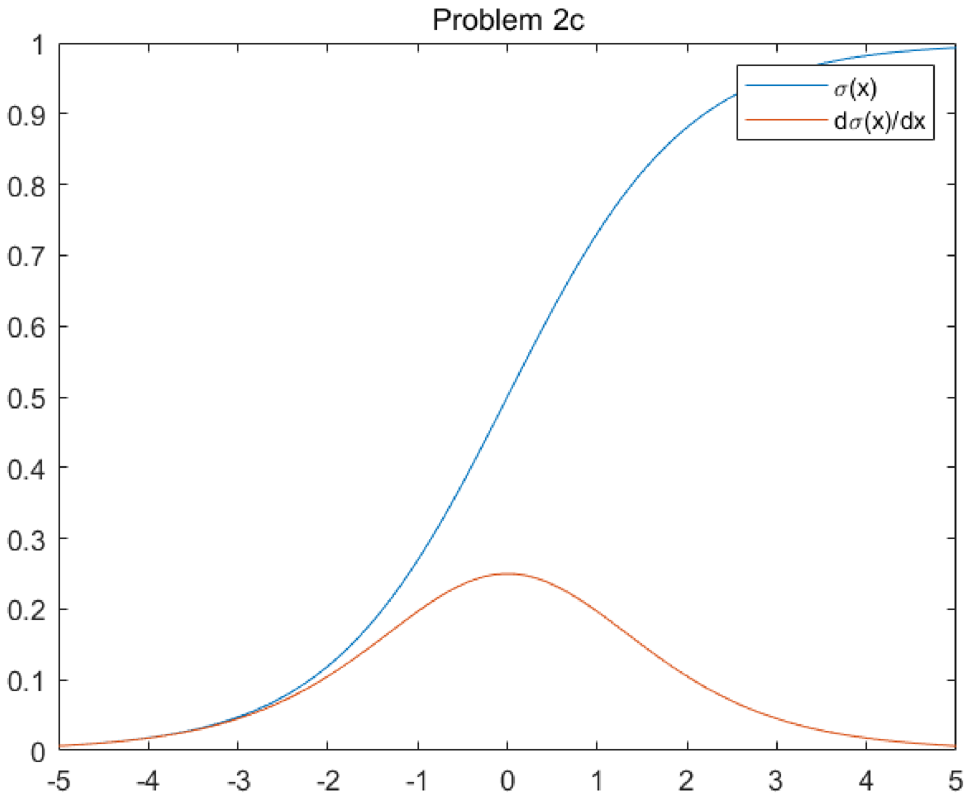

Figure 1: problem 2c

1.3

$$
\begin{aligned}
f(c) &= \frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 + c^2 - 2x_i \cdot c\right) \\
&= c^2 - \frac{2}{n}\sum_{i=1}^{n}(x_i \cdot c) + \frac{1}{n}\sum_{i=1}^{n}(x_i^2) \\
&= (c - \frac{1}{n}\sum_{i=1}^{n} x_i)^2 + \frac{1}{n}\sum_{i=1}^{n}(x_i^2) - (\frac{1}{n}\sum_{i=1}^{n}(x_i))^2
\end{aligned}
$$

Since x_i is fixed, the minimum c is when $c = \frac{1}{n}\sum_{i=1}^{n} x_i$

1.4

From this problem, we let $A =$ who has cancer, $B =$ who is positive in tests
It is a conditional probability problem, and we need to know $P(A|B)$
From Bayes' theorem,

$$
\begin{aligned}
P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} &= \frac{P[B \mid A] \cdot P[A]}{P[B \mid A] \cdot P[A] + P[B \mid \neg A] \cdot P[\neg A]} \\
&= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.096 \times 0.99} = 0.0777
\end{aligned}
$$

1.5

We should choose $b, 6.1$
Since it is almost a normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and

$$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

Since $\mu = 150, f(\mu) = 0.063$, we can get that

$$\sigma = \frac{1}{f(\mu)\sqrt{2\pi}} = 6.3324 \approx 6.1$$