Northeastern University
College of Engineering
Group member name(s): Yanming Liu
Group member UID(s): 002199402

June 5, 2022
Homework 1

IE 7374: Machine Learning

Answer to Question 1

a)
$E(x) = \sum_{x=1}^{4} xp(x) = 1 \cdot \frac{1}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{3}{10} + 4 \cdot \frac{4}{10} = 3$

$Var(x) = E(x^2) - [E(x)]^2 = 10 - 9 = 1$, where $E(x^2) = \sum_{x=1}^{4} x^2 p(x) = 10$

b)
$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$
$1)y = 0$, $p(y) = \sum_{x=1}^{4} p(y=0|x)p(x) = \frac{1}{4} \cdot \frac{2}{10} + \frac{1}{3} \cdot \frac{3}{10} + \frac{3}{8} \cdot \frac{4}{10} = \frac{3}{10}$

- $x = 1$, $p(y=0|x=1) = 0$, $p(x=1) = \frac{1}{10}$, $p(x=1|y=0) = 0$
- $x = 2$, $p(y=0|x=2) = \frac{1}{4}$, $p(x=1) = \frac{2}{10}$, $p(x=2|y=0) = \frac{1}{4} \cdot \frac{2}{10} / \frac{3}{10} = \frac{1}{6}$
- $x = 3$,$p(y=0|x=3) = \frac{1}{3}$, $p(x=3) = \frac{3}{10}$, $p(x=3|y=0) = \frac{1}{3}$
- $x = 4$,$p(y=0|x=4) = \frac{3}{8}$, $p(x=4) = \frac{4}{10}$, $p(x=4|y=0) = \frac{1}{2}$

Similarly, $y = 1$

- $p(x=1|y=1) = \frac{1}{7}$
- $p(x=2|y=1) = \frac{3}{14}$
- $p(x=3|y=1) = \frac{2}{7}$
- $p(x=4|y=1) = \frac{5}{14}$

c)
$E[X|Y=1] = \sum_{x=1}^{4} p(x|y=1)p(x) = \frac{1}{7} \cdot 1 + \frac{3}{14} \cdot 2 + \frac{2}{7} \cdot 3 + \frac{5}{14} \cdot 4 = \frac{20}{7}$

$$
\begin{aligned}
\theta_{MAP} &= argmax \prod_{i=1}^{N} P(y_i|\theta^T x_i, \sigma^2)P(\theta) \\
&= argmax \prod_{i=1}^{N} [\frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(y_i - \theta^T x_i)^2}{2\sigma^2} \cdot \frac{\lambda}{2} \exp^{-\lambda|\theta|}] \\
&\Longrightarrow argmax \sum_{i=1}^{N} [\frac{-(y_i - \theta^T x_i)^2}{2\sigma^2} - \lambda|\theta|]
\end{aligned}
$$

Answer to Question 2

$$
\begin{aligned}
Cov(x,y) &= E[(x-E[x])(y-E[y])] \\
&= E[xy - xE[y] - yE[x] + E[x]E[y]] \\
&= E[xy] - E[xE[y]] - E[yE[x]] + E[x]E[y] \\
&= E[x][y] - E[y]E[x] - E[x]E[y] + E[x]E[y] \\
&= 0
\end{aligned}
$$

The prove holds trues only when $E[xy] = E[x]E[y]$, which means variables x, y are independent.

Answer to Question 3

a)
$E[f[X] = \sum_{x_i \in \{a,b,c\}} p(f(x_i))p(x_i) = 10 \cdot 0.1 + 5 \cdot 0.2 + \frac{10}{7} \cdot 0.7 = 3$

b)
$E[1/p(X)] = \sum_{x_i \in \{a,b,c\}} \frac{1}{p(x_i)} p(x_i) = 3$

Answer to Question 4

a)
$X = \{X_1, X_2, \ldots, X_i, \ldots X_n\},\ X_i \sim \mathcal{N}(\mu, \sigma^2),\ \theta = (\mu, \sigma^2)$
For MLE,

$$
\begin{aligned}
P = P(X|\theta) \quad &= \quad \prod_{i=1}^{n} P(X_i|\theta) \\
&\xrightarrow{Apply\ log} \sum_{i=1}^{n} logP(X_i|\theta) \\
&= \quad \sum_{i=1}^{n} log(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \frac{-(X_i-\mu)^2}{2\sigma^2}) \\
&= \quad \sum_{i=1}^{n} [log\frac{1}{\sqrt{2\pi}} - log\sigma - \frac{(X_i-\mu)^2}{2\sigma^2}]
\end{aligned}
$$

Find the partial deviates of μ and σ^2,

$$\frac{\partial P}{\partial \mu} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma^2}$$

$$let \frac{\partial P}{\partial \sigma} = \sum_{i=1}^{n} [\frac{1}{\sigma} + \frac{1}{2}(X_i-\mu)^2 \cdot (-2)\sigma^{-3}]$$

let $\frac{\partial P}{\partial \mu} = 0,\ \frac{\partial P}{\partial \sigma} = 0,$

$$\hat{\mu} = \frac{1}{n}\sum_{i_1}^{n} X_i$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

b)
$E[\hat{\mu}] = \mu$, unbiased; $E[\hat{\sigma^2}] = \frac{n-1}{n}\sigma^2$, biased
Prove:
To verify if biased, We test the expectation of respective predicting parameters in sample data using MLEs with these real parameters under normal distribution having overall data records. If these values are equal, then biased, otherwise, unbiased.

$$E[\hat{\mu}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$

$$\begin{aligned}
E[\hat{\sigma^2}] &= E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2] \\
&= E[\frac{1}{n}\sum_{i}^{n}[(X_i - \mu) - (\hat{\mu} - \mu)]^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{2}{n}\sum_{i=1}^{n}(X_i - \mu)(\hat{\mu} - \mu) + \frac{1}{n}\sum_{i}^{n}(\hat{\mu} - \mu)^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2(\hat{\mu} - \mu)^2 + (\hat{\mu} - \mu)^2] \\
&= E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\hat{\mu} - \mu)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i - \mu)^2 - E[(\hat{\mu} - \mu)^2] \\
&= \sigma^2 - Var[\hat{\mu}] \\
&= \sigma^2 - \frac{1}{n}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

An unbiased for σ^2 can be obtained, $\sigma_{ub} = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$

Answer to Question 5

Suppose variable X and Y are continuous one,

$$\begin{aligned}
E[XY] &= \iint_{x,y} xy \cdot p(xy)dxdy \\
&= \iint_{x,y} xy \cdot p(x)p(y)dxdy \\
&= \int_x xp(x)dx \cdot \int_y yp(y)dy \\
&= E[x]E[y]
\end{aligned}$$

$p(xy) = p(x)p(y)$ holds if variables x, y are independent.

Answer to Question 6

a) ii

$$P(H|e1,e2) = \frac{P(e1,e2|H)P(H)}{P(e1,e2)}$$

b) i & ii

$$\begin{aligned}P(H|e1,e2) &= \frac{P(e1,e2|H)P(H)}{P(e1,e2)} \\ &= \frac{P(e1|H)P(e2|H)P(H)}{P(e1,e2)}\end{aligned}$$

Answer to Question 7

$$\begin{aligned}Var[X+Y] &= E[(X+Y-E[X+Y])^2] \\ &= E[(X+Y)^2 - 2(X+Y)E[X+Y] + (E[X+Y])^2)] \\ &= E[X^2] + 2E[XY] + E[Y^2] - 2E[X+Y]E[X+Y] \\ &\quad + (E[X])^2 + (E[Y])^2 + 2E[X]E[Y]) \\ &= E[X^2] + 2E[XY] + E[Y^2] - 2[(E[X])^2 + 2E[X]E[Y] + (E[Y])^2] \\ &\quad + (E[X])^2 + (E[Y])^2 + 2E[X]E[Y]) \\ &= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2(E[XY] - E[X]E[Y]) \\ &= Var[X] + Var[Y] + 2Cov[X,Y]\end{aligned}$$

If variables X, Y are independent, $Cov[X,Y] = 0$, according the insights from Question 5. Then, $Var[X+Y] = Var[X] + Var[Y]$

Answer to Question 8

We can easily tell that $p(x|\theta_0)$ is an increasing function with respect to θ_0. This can be obtain by calculating the derivative of joint probability function, $\frac{\partial logp(x|\theta_0)}{\theta_0} = n > 0$. It shows $logp(x|\theta_0)$ is an increasing function, and $log(x)$ is also an increasing function. Then, $p(x|\theta_0)$ increase, too. As $\theta_0 <= x$, the maximum like-hood estimate of θ_0 is $min(x_1, x_2, \ldots, x_n)$.

Answer to Question 9

a)
To prove a matrix A is a positive-semi definite, which is to show $x^T Ax >= 0$.

$$x^T(A^T A)x = (Ax)^T Ax = \|Ax\|^2 >= 0$$

b)
If A is full rank, Rank(A) = n & $m >= n$, then $A^T A$ is positive definite.