CS57300: Homework 1 - SOLUTION

Your homework must be typed and submitted as a PDF. Use of Latex is recommended, but not required.

1 Counting (5 pts)

(a) Two cards are drawn from a deck of 52 cards without replacement:

(i) What is the probability that the second card is a heart, given that the first card is a heart?

$$\frac{13-1}{51} = \frac{12}{51}$$

(ii) What is the probability that none of the cards are hearts, given that at most one card is a heart?

$$P(\text{None of the cards are hearts}|\text{At most one card is heart})$$

$$= \frac{P(\text{None of cards are hearts, At most one card is heart})}{P(\text{At most one card is heart})} = \frac{\frac{39}{52} \times \frac{38}{51}}{1 - \frac{13}{52} \times \frac{12}{51}} = 0.59375$$

(b) One card is selected from a deck of 52 cards and placed in a second deck containing 52 cards. A card is then selected from the second deck.

(i) What is the probability that a card drawn from the second deck is an ace?

$$\frac{4}{52} \times \frac{52}{53} + \frac{4}{52} \times \frac{1}{53} = \frac{4}{52} = 0.07692$$

(ii) If the first card is placed into a deck of 54 cards containing two jokers, then what is the probability that a card drawn from the second deck is an ace?

$$\frac{4}{52} \times \frac{1}{55} + \frac{4}{54} \times \frac{54}{55} = 0.074125$$

(iii) Given that an ace was drawn from the second deck in (ii), what is the conditional probability that an ace was transferred from the first deck?

$$= P(\text{Ace transferred from first deck}|\text{Ace is drawn from second deck})$$

$$= \frac{P(\text{Ace transferred from first deck, Ace is drawn from second deck})}{P(\text{Ace is drawn from second deck})}$$

$$= \frac{\frac{4}{52} \times \frac{5}{55}}{\frac{4}{52} \times \frac{1}{55} + \frac{4}{54} \times \frac{54}{55}} = 0.094339$$

2 Probability and conditional probability (4 pts)

(a) Suppose that 30 percent of computer owners use an Apple machine, 50 percent use a Windows machine, and 20 percent use Linux. Suppose that 40 percent of Apple users have succumbed to a computer virus, 76 percent of Windows users get the virus, and 55 percent of Linux users get the virus. We select a person at random and learn that their system was infected with the virus. What is the probability that the person is a Windows user?

$$= P(\text{Windows user}|\text{Has virus}) = \frac{P(\text{Windows user, Has virus})}{P(\text{Has virus})}$$

$$= \frac{\frac{50}{100} \times \frac{76}{100}}{\frac{30}{100} \times \frac{40}{100} + \frac{50}{100} \times \frac{76}{100} + \frac{20}{100} \times \frac{55}{100}} = 0.62295$$

(b) There are three cards. The first is green on both sides, the second is red on both sides, and the third is green on one side and red on the other. Consider the scenario where a card is chosen at random and one side is shown (also chosen at random). If the side shown is green, what is the probability that the other side is also green?

$$= P(\text{back=Green}|\text{front=Green})$$

- Card 1: Side A = G, Side B = G
- Card 2: Side A = R, Side B = R
- Card 3: Side A = R, Side B = G

If we know the front is green, it is either Card1 Side A, or Card1 Side B, or Card 3 Side B. So, the answer is $\frac{2}{3}$.

3 Probability distributions (5 pts)

(a) Let X be a random variable with discrete pdf $f(x) = \frac{x}{8}$ if $x = 1, 2, \text{or } 5$ and zero otherwise.

(i) Sketch the graph of the discrete pdf $f(x)$.
Please see Figure 1.

(ii) Find $E[X]$ and $Var(X)$.

$$E[X] = \sum_x x.P(x) = 1.\frac{1}{8} + 2\frac{2}{8} + 5\frac{5}{8} = 3.75$$

$$Var[X] = \sum_x (x - E[X])^2.P(x) = (1 - 3.75)^2\frac{1}{8} + (2 - 3.75)^2\frac{2}{8} + (5 - 3.75)^2\frac{5}{8}$$
$$= 2.6875$$

(iii) Find $E[2X + 3]$

$$E[2X + 3] = E[2X] + 3 = 2E[X] + 3 = 2 \times 3.75 + 3 = 10.5$$

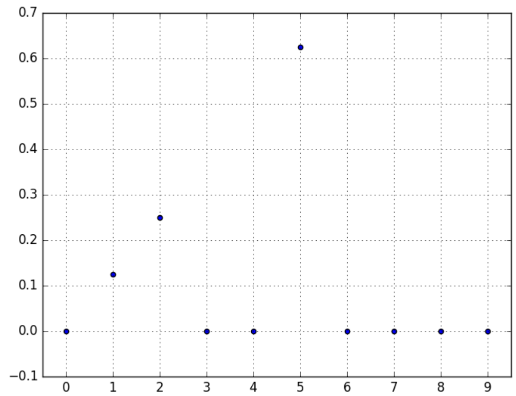

Figure 1: PMF

(b) The form of the Bernoulli(p) distribution is not symmetric between the two values of X. In some situations, it will be more convenient to use an equivalent formulation for which $x \in \{-1, 1\}$, in which case the distribution can be written as:

$$P(x|p) = \left(\frac{1-p}{2}\right)^{(1-x)/2} \left(\frac{1+p}{2}\right)^{(1+x)/2}$$

Show that this distribution is normalized (i.e., sums to 1) and evaluate its mean and variance.

$$\begin{aligned}\sum_{x\in\{-1,1\}} P(x|p) &= P(x=-1|p) + P(x=1|p) \\ &= \left(\frac{1-p}{2}\right) \times \left(\frac{1+p}{2}\right)^0 + \left(\frac{1-p}{2}\right)^0 \times \left(\frac{1+p}{2}\right) \\ &= \left(\frac{1-p}{2}\right) + \left(\frac{1+p}{2}\right) = \frac{2}{2} = 1\end{aligned}$$

$$E[X] = \sum_{x\in\{-1,1\}} x.P(x) = -1 \times \left(\frac{1-p}{2}\right) + 1 \times \left(\frac{1+p}{2}\right) = p$$

$$\begin{aligned}Var[X] = \sum_{x\in\{-1,1\}} (x - E[X])^2 P(x) &= (-1-p)^2 \times \left(\frac{1-p}{2}\right) + (1-p)^2 \times \left(\frac{1+p}{2}\right) \\ &= (1+2p+p^2) \times \left(\frac{1-p}{2}\right) + (1-2p+p^2) \times \left(\frac{1+p}{2}\right) \\ &= \frac{(1+2p+p^2-p-2p^2-p^3) + (1-2p+p^2+p-2p^2+p^3)}{2} \\ &= \frac{2-2p^2}{2} = 1-p^2\end{aligned}$$

4 Independence (5 pts)

(a) Prove the following:
If A and B are independent events, then $P(A|B) = P(A)$.
Also, for any pair of events A and B,

$$P(AB) = P(A|B)P(B) = P(B|A)P(A)$$

The definition of independent events:

$$P(A, B) = P(A)P(B)$$

We know that for any pair of events A and B,

$$P(A, B) = P(A|B)P(B)$$

So, we can:

$$P(A, B) = P(A)P(B) = P(A|B)P(B)$$

Finally:

$$P(A) = P(A|B)$$

(b) Prove the following:
If A and B are conditionally independent given Z, that is, $P(A, B|Z) = P(A|Z)P(B|Z)$, then $P(A|B, Z) = P(A|Z)$.
The definition of conditional independence is:

$$P(A, B|Z) = P(A|Z)P(B|Z)$$

The definition of conditional probability gives:

$$P(A, B|Z) = P(A, B, Z)/P(Z)$$

$$P(A|Z) = P(A, Z)/P(Z)$$

$$P(B|Z) = P(B, Z)/P(Z)$$

So, we know:

$$P(A, B, Z) = P(A, Z)P(B, Z)/P(Z)$$

Moving $P(B, Z)$ to the lefthand side:

$$P(A, B, Z)/P(B, Z) = P(A, Z)/P(Z)$$

Given the definition of conditional probability, we have:

$$P(A|B, Z) = P(A|Z)$$

(c) A box contains the following four slips of paper, each having exactly the same dimensions: (1) `win prize 1`, (2) `win prize 2`, (3) `win prize 3`, (4) `win prizes 1, 2, and 3`. One slip will be randomly selected. Let A_1 = `win prize 1`, A_2 = `win prize 2`, and A_3 = `win prize 3`. Show that A_1, A_2, and A_3 are *pairwise* independent, but that the three events are not *mutually* independent (i.e., $P(A_1 \wedge A_2 \wedge A_3) \neq P(A_1)P(A_2)P(A_3)$).

- $P(A_1) = P(A_2) = P(A_3) = \frac{2}{4}$
- $P(A_1, A_2) = P(A_1, A_3) = P(A_2, A_3) = \frac{1}{4}$
- $P(A_1, A_2, A_3) = \frac{1}{4}$
- $P(A_1, A_2) = \frac{1}{4} = P(A_1)P(A_2) = \frac{1}{2} \times \frac{1}{2}$ (Correct)
- $P(A_1, A_3) = \frac{1}{4} = P(A_1)P(A_3) = \frac{1}{2} \times \frac{1}{2}$ (Correct)
- $P(A_2, A_3) = \frac{1}{4} = P(A_2)P(A_3) = \frac{1}{2} \times \frac{1}{2}$ (Correct)
- Because of the last three bullet points are correct, we can conclude that A_1, A_2, and A_3 are *pairwise* independent.
- $P(A_1, A_2, A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$
- Because of the last bullet point, we can conclude that A_1, A_2, and A_3 are not *mutually* independent.

5 Expectation (5 pts)

(a) Let $X_1, ..., X_n \sim$ Bernoulli(p=0.5). Let $Y_n = \max\{X_1, ..., X_n\}$.

(i) Find $E[Y_n]$.

- The probability of 0 success in n trials: $(1-p)^n$
- The probability of at least 1 success in n trials: $1-(1-p)^n$
- Probability mass function:

$$Y_n = \begin{cases} (1-p)^n & x = 0 \\ 1-(1-p)^n & x = 1 \end{cases}$$

$$E[Y_n] = 0.(1-p)^n + 1.(1-(1-p)^n) = 1-(1-p)^n = 1 - 0.5^n$$

(ii) Plot $E[Y_n]$ as a function of n.

Please see Figure 2.

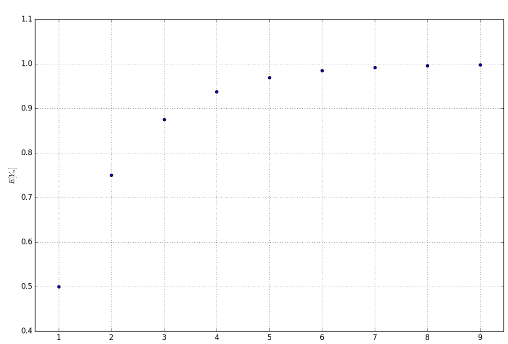

Figure 2: $E[Y_n]$ as a function of n.

(iii) How is the distribution of the max (Y_n) different from that of a single Bernoulli (X_i)?

The expectation of the Bernoulli is constant. On the other hand, the expectation of max converges to 1.

Also, the distribution of max depends on n whereas the distribution of individual X_n does not depend on n.

(b) You and your friend are playing the following game: two dice are rolled; if the total showing is divisible by 4, you pay your friend \$12. If you want to make the game fair, how much should she pay you when the total is not divisible by 4? A fair game is one in which your expected winnings are \$0.

4, 8 and 12 are divisible by four. The total probability:

$$P(T = 4|T = 8|T = 12) = 3/36 + 5/36 + 1/36 = 9/36 = 1/4$$

So, in other words:

$$P(\text{Total is divisible by four}) = 1/4$$

And,

$$P(\text{Total is not divisible by four}) = 1 - 1/4 = 3/4$$

By definition:

$$0 = P(\text{Total is divisible by four}).(-12) + P(\text{Total is not divisible by four}).R$$

We get:

$$R = 4$$

Thus, to have a fair game, she should pay us \$4 when the total is not divisible by 4.

6 Conditional Expectation (4 pts)

Consider the setting where you first roll a fair 6-sided die, and then you flip a fair coin the number of times shown by the die. Let D refer to the outcome of the die roll (i.e., number of coin flips) and let H refer to the number of heads observed after D coin flips.

(a) Determine $E[H|d]$ and $Var(H|d)$.

$$P(H = h|d) = \binom{d}{h}(p)^h.(1-p)^{(d-h)}$$

where $p = 0.5$.

We see that this follows a Binomial distribution, which has the properties:

$$E[X] = np$$

$$Var[X] = np(1-p)$$

So, expected values:

$$\begin{aligned}E[H|d=1] &= 0.5\\ E[H|d=2] &= 1\\ E[H|d=3] &= 1.5\\ E[H|d=4] &= 2\\ E[H|d=5] &= 2.5\\ E[H|d=6] &= 3\end{aligned}$$

And variance:

$$\begin{aligned}Var[H|d=1] &= 10.5(1-0.5) = 0.25\\ Var[H|d=2] &= 2.0.5(1-0.5) = 0.5\\ Var[H|d=3] &= 3.0.5(1-0.5) = 0.75\\ Var[H|d=4] &= 4.0.5(1-0.5) = 1\\ Var[H|d=5] &= 5.0.5(1-0.5) = 1.25\\ Var[H|d=6] &= 6.0.5(1-0.5) = 1.5\end{aligned}$$

(b) Determine $E[H]$ and $Var(H)$

$$\begin{aligned}E[H] &= \sum_d E[H|D=d]P(D=d)\\ &= 0.5.(1/6) + 1.(1/6) + 1.5.(1/6) + 2.(1/6) + 2.5.(1/6) + 3.(1/6)\\ &= 1.75\end{aligned}$$

$$\begin{aligned}Var(E[H|D]) &= ((0.5-1.75)^2 + (1-1.75)^2 + (1.5-1.75)^2 + (2-1.75)^2 + (2.5-1.75)^2 + (3-1.75)^2)/6\\ &= 0.72916\end{aligned}$$

$$\begin{aligned}Var(H) &= E[Var(H|D)] + Var(E[H|D])\\ &= (1.75/2) + 0.72916\\ &= 1.60416\end{aligned}$$

7 Covariance and Correlation (6 pts)

(a) Show that if $E[X|Y=y] = c$ for some constant c, then X and Y are uncorrelated.

Assuming X and Y are continuous random variables,

$$\begin{aligned}E[X] &= \int E[X|Y=y]\ p(y)\ dy = c\int p(y)\ dy = c\\ E[XY] &= \iint x\ y\ p(x,y)\ dx\ dy = \iint xy\ p(x|y)p(y)\ dx\ dy\\ &= \int\left(\int x\ p(x|y)\ dx\right) y\ p(y)\ dy = \int E[X|Y=y]\ y\ p(y)\ dy\\ &= c\int y\ p(y)\ dy = cE[Y]\end{aligned}$$

So $cov(X,Y) = E[XY] - E[X]E[Y] = cE[Y] - cE[Y] = 0$. Since $Cov(X,Y) = 0$, we can conclude that X and Y are uncorrelated. Because:

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_x.\sigma_y}$$

If $Cov(X,Y) = 0$, then $Cor(X,Y) = 0$.

(b) Show $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$.

Using the definition of Covariance:

$$\begin{aligned}
Cov(X,Y+Z) &= E[(Z+Y-E[Z+Y])(X-E[X])] \\
&= E[(Z+Y-E[Z]-E[Y])(X-E[X])] \\
&= E[((Z-E[Z])+(Y-E[Y]))(X-E[X])] \\
&= E[(Z-E[Z])(X-E[X])+(Y-E[Y])(X-E[X])] \\
&= E[(Z-E[Z])(X-E[X])]+E[(Y-E[Y])(X-E[X])] \\
&= Cov(X,Z)+Cov(X,Y) \\
&= Cov(X,Y)+Cov(X,Z)
\end{aligned}$$

(c) Let X_1 and X_2 be quantitative and verbal scores on one aptitude exam and Y_1 and Y_2 be corresponding scores on another exam. If $Cov(X_1, Y_1) = 5$, $Cov(X_1, Y_2) = 1$, $Cov(X_2, Y_1) = 2$, and $Cov(X_2, Y_2) = 8$, what is the covariance between the two total scores $X_1 + X_2$ and $Y_1 + Y_2$?

Bilinearity (see 7.b.) and Symmetry principles of Covariance:

$$\begin{aligned}
Cov(X_1+X_2, Y) &= Cov(X_1, Y) + Cov(X_2, Y) \\
Cov(X,Y) &= Cov(Y,X)
\end{aligned}$$

Using those:

$$\begin{aligned}
Cov(X_1+X_2, Y_1+Y_2) &= Cov(X_1, Y_1+Y_2) + Cov(X_2, Y_1+Y_2) \\
&= Cov(X_1,Y_1) + Cov(X_1,Y_2) + Cov(X_2,Y_1) + Cov(X_2,Y_2) \\
&= 5+1+2+8 \\
&= 16
\end{aligned}$$

8 Distance and Correlation Measures (5 pts)

(a) Show how Euclidean distance can be expressed as a function of **cosine similarity** when each data vector has an L_2 length of 1.

Let's assume,

$$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad \text{for some} \quad n > 0$$

L_2 norm:

$$L_2(\mathbf{x}) = ||\mathbf{x}|| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Euclidean distance:

$$d(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}|| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Cosine similarity:

$$\text{cos\_sim}(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}\cdot\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

We know that $||\mathbf{x}|| = 1, ||\mathbf{y}|| = 1$.

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &= ||\mathbf{x}-\mathbf{y}|| = \sqrt{(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})} = \sqrt{\mathbf{x}\cdot\mathbf{x} - 2\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{y}} \\ &= \sqrt{||\mathbf{x}||^2 - 2\mathbf{x}\cdot\mathbf{y} + ||\mathbf{y}||^2} = \sqrt{1 - 2\mathbf{x}\cdot\mathbf{y} + 1} \\ &= \sqrt{2 - 2\mathbf{x}\cdot\mathbf{y}} = \sqrt{2 - 2\text{cos\_sim}(\mathbf{x},\mathbf{y})} \end{aligned}$$

(b) Show how Euclidean distance can be expressed as a function of **correlation** when each data point has been standardized by subtracting its mean and dividing by its standard deviation.

Correlation:

$$Cor(\mathbf{x},\mathbf{y}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$$

We know $\mu_x = 0, \mu_y = 0, \sigma_{\mathbf{x}} = 1, \sigma_{\mathbf{y}} = 1$:

$$Cor(\mathbf{x},\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{n}\mathbf{x}\cdot\mathbf{y}$$

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &= ||\mathbf{x}-\mathbf{y}|| = \sqrt{(\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})} = \sqrt{\mathbf{x}\cdot\mathbf{x} - 2\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{y}} \\ &= \sqrt{n - 2\mathbf{x}\cdot\mathbf{y} + n} = \sqrt{2n - 2\mathbf{x}\cdot\mathbf{y}} \\ &= \sqrt{2n - 2nCor(\mathbf{x},\mathbf{y})} = \sqrt{2n(1 - Cor(\mathbf{x},\mathbf{y}))} \end{aligned}$$

Note: We know that $\mathbf{x}\cdot\mathbf{x} = n$ and $\mathbf{y}\cdot\mathbf{y} = n$ because:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} = \sqrt{\frac{1}{N}\mathbf{x}\cdot\mathbf{x}}$$

$$\sigma_x^2 = \frac{1}{N}\mathbf{x}\cdot\mathbf{x}$$

We know $\sigma_x = 1$, thus:

$$\mathbf{x}\cdot\mathbf{x} = n$$

Notice that same argument applies to $\mathbf{y}\cdot\mathbf{y} = n$.

9 Linear Algebra (2 pts)

(a) Specify whether the following matrix has an inverse without trying to compute the inverse: (*Recall that a matrix A is invertible if and only if* $\det(A) \neq 0$.)

$$\begin{bmatrix} 9 & 1 & 9 & 9 & 9 \\ 9 & 0 & 9 & 9 & 2 \\ 4 & 0 & 0 & 5 & 0 \\ 9 & 0 & 3 & 9 & 0 \\ 6 & 0 & 0 & 7 & 0 \end{bmatrix}$$

$$\begin{vmatrix} 9 & 1 & 9 & 9 & 9 \\ 9 & 0 & 9 & 9 & 2 \\ 4 & 0 & 0 & 5 & 0 \\ 9 & 0 & 3 & 9 & 0 \\ 6 & 0 & 0 & 7 & 0 \end{vmatrix} = -1 \times \begin{vmatrix} 9 & 9 & 9 & 2 \\ 4 & 0 & 5 & 0 \\ 9 & 3 & 9 & 0 \\ 6 & 0 & 7 & 0 \end{vmatrix} = -1 \times -2 \times \begin{vmatrix} 4 & 0 & 5 \\ 9 & 3 & 9 \\ 6 & 0 & 7 \end{vmatrix}$$

$$= 2 \times \left(3 \times \begin{vmatrix} 4 & 5 \\ 6 & 7 \end{vmatrix} \right) = 2 \times (3 \times (4 \times 7 - 5 \times 6)) = -12$$

Since $\det(A) \neq 0$, we conclude that the matrix A is invertible.

(b) Find eigenvalues and eigenvectors of the following matrix:

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} -\lambda & 1 \\ -2 & -3 - \lambda \end{bmatrix}$$

Set the determinant of $A - \lambda I = 0$:

$$-\lambda(-3 - \lambda) - 1 \times (-2) = \lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2) = 0$$

The two eigenvalues are thus:

$$\lambda_1 = -1, \lambda_2 = -2$$

Now let's find the eigenvectors associated with the first eigenvalue $\lambda_1 = -1$. We have:

$$(A - \lambda_1 I)x = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} x = 0$$

$$x_1 = \begin{bmatrix} k_1 \\ -k_1 \end{bmatrix}$$

For the eigenvectors associated with the second eigenvalue $\lambda_2 = -2$, we have:

$$(A - \lambda_2 I)x = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} x = 0$$

$$x_2 = \begin{bmatrix} k_2 \\ -2k_2 \end{bmatrix}$$

10 Statistical Inference (6 pts)

Suppose you obtain N data points $X = \{x_1, x_2, ..., x_N\}$ from a normal distribution whose variance is δ^2 and mean is unknown.

(a) What is the maximum likelihood estimation of the normal distribution's mean value μ?

The log likelihood of observing X is:

$$logP(X|\mu) = log\prod_{i=1}^{N}\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\delta^2}}$$

$$= \sum_{i=1}^{N} log(\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\delta^2}})$$

$$= -Nlog\delta - \frac{N}{2}log(2\pi) - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\delta^2}$$

To maximize the likelihood, we have:

$$\frac{dlogP(X|\mu)}{d\mu} = \sum_{i=1}^{N}\frac{x_i-\mu}{\delta^2} = 0$$

$$\mu_{MLE} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

(b) If the prior distribution for μ is a normal distribution with mean value of η and variance of λ^2, i.e., $\mu \sim N(\eta, \lambda^2)$, what is the maximum a-posteriori estimation of μ?

The posterior distribution of μ is:

$$P(\mu|X) \propto P(\mu)P(X|\mu) = \frac{1}{\lambda\sqrt{2\pi}}e^{-\frac{(\mu-\eta)^2}{2\lambda^2}}\prod_{i=1}^{N}\frac{1}{\delta\sqrt{2\pi}}e^{-\frac{(x_i-\mu)^2}{2\delta^2}}$$

Take the log on both sides:

$$logP(\mu|X) \propto -log\lambda - Nlog\delta - \frac{N+1}{2}log(2\pi) - \frac{(\mu-\eta)^2}{2\lambda^2} - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\delta^2}$$

To get the mode of the posterior distribution, let's set the first derivative to zero:

$$-\frac{\mu-\eta}{\lambda^2} + \sum_{i=1}^{N}\frac{x_i-\mu}{\delta^2} = 0$$

Therefore:

$$\mu_{MAP} = \frac{\sum_{i=1}^{N}\frac{x_i}{\delta^2} + \frac{\eta}{\lambda^2}}{\frac{N}{\delta^2} + \frac{1}{\lambda^2}}$$