Summer 2020 CX4641/CS7641 Homework 1

Dr. Mahdi Roozbahani

Deadline: May 29, Friday, 11:59 pm

- No extension of the deadline is allowed. Late submission will lead to 0 credit.
- Discussion is encouraged on Piazza as part of the Q/A. However, all assignments should be done individually.

Instructions

- This assignment has no programming, only written questions.
- We will be using Gradescope this semester for submission and grading of assignments.
- Your write up must be submitted in PDF form, you may use either Latex or markdown, whichever you prefer. We will not accept handwritten work.
- Please make sure to start answering each question on a new page. It makes it more organized to map your answers on GradeScope. When submitting your assignment, you must correctly map pages of your PDF to each question/subquestion to reflect where they appear. Improperly mapped questions may not be graded correctly.

1 Linear Algebra [25pts + 8pts]

1.1 Determinant and Inverse of Matrix [11pts]

Given a matrix M:

$$M = \begin{bmatrix} 2 & -1 & 1 \\ 4 & 1 & -2 \\ 2 & -1 & 3 \end{bmatrix}$$

(a) Calculate the determinant of M. [2pts] (Calculation process required.)

(b) Calculate M^{-1}. [5pts] (Calculation process required)
(**Hint:** Please double check your answer and make sure $MM^{-1} = I$)

(c) What is the relationship between the determinant of M and the determinant of M^{-1}? [2pts]

(d) When does a matrix not have an inverse? Provide an example. [2pts]

Solution:

(a) Determinant of M is given by:

$$det(M) = 2[(1 \times 3) - (-2 \times -1)] - (-1)[(4 \times 3) - (2 \times -2)] + 1[(4 \times -1) - (1 \times 2)]$$

$$\Rightarrow \boxed{det(M) = 12}$$

(b) The inverse of matrix M is given by:

$$M^{-1} = \frac{adj(M)}{det(M)}$$

where $adj(M)$ represents the adjugate of M (or transpose of cofactor matrix).
The cofactor matrix is given by:

$$C = \begin{bmatrix} [(1 \times 3) - (-1 \times -2)] & -[(4 \times 3) - (2 \times -2)] & (4 \times -1) - (1 \times 2) \\ -[(-1 \times 3) - (-1 \times 1)] & [(2 \times 3) - (2 \times 1)] & -[(2 \times -1) - (2 \times -1)] \\ [(-1 \times -2) - (1 \times 1)] & -[(2 \times -2) - (4 \times 1)] & [(2 \times 1) - (4 \times -1)] \end{bmatrix}$$

So, we have:

$$adj(M) = C^T = \begin{bmatrix} 1 & 2 & 1 \\ -16 & 4 & 8 \\ -6 & 0 & 6 \end{bmatrix}$$

$$\Rightarrow \boxed{M^{-1} = \frac{1}{12} \begin{bmatrix} 1 & 2 & 1 \\ -16 & 4 & 8 \\ -6 & 0 & 6 \end{bmatrix}}$$

(c) The relationship between the determinant of a matrix M and its inverse is given by:

$$\boxed{det(M) = \frac{1}{det(M^{-1})}}$$

(d) A matrix does not have an inverse when it is singular (i.e. determinant is zero). An example singular matrix Z is:

$$\boxed{Z = \begin{bmatrix} 1 & 2 & 1 \\ -5 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}}$$

1.2 Characteristic Equation [8pts] (Bonus)

Consider the eigenvalue problem:

$$Ax = \lambda x, x \neq 0$$

where x is a non-zero eigenvector and λ is eigenvalue of A. Prove that the determinant $|A - \lambda I| = 0$.

Solution:

From the eigenvalue problem, it follows that:

$$(A - \lambda I) \cdot x = 0$$

where 0 indicates a null matrix and I is the identity matrix with the same dimension as A.
We can now represent the above equation as a generalized linear system

$$Mx = 0$$

From the matrix inverse relation below, we get that:

$$M^{-1} = \frac{adj(M)}{det(M)} \Rightarrow M^{-1}M = I_n = \frac{adj(M)}{det(M)}M \Rightarrow adj(M)M = I_n det(M)$$

where I_n is the identity matrix of the same dimension as M. Now, as long as $adj(M)$ is not a null matrix (trivial solution), we can say that:

$$adj(M)Mx = 0 \Rightarrow det(M) \cdot I_n x = 0$$

Thus, the linear system $M \cdot x = 0$ has a non-trivial solution iff M is singular. So, we have:

$$\boxed{det(A - \lambda I) = 0}$$

1.3 Singular Value Decomposition [14pts]

Given a matrix A:

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

Compute the Singular Value Decomposition (SVD) by following the steps below. Your full calculation process is required.

(a) Calculate all eigenvalues of AA^T and A^TA. The square roots of the positive eigenvalues make up the singular values, the diagonal entries in Σ. They will be arranged in descending order, all other values in Σ are 0. [4pts]

(b) Calculate all eigenvectors of AA^T normalized to unit length. These will make up the left singular vectors, or the columns of U. [4pts]

(c) Calculate all eigenvectors of A^TA normalized to unit length. These will make up the right singular vectors, or the rows of V^T. [4pts]

(d) Put it all together. Write out the SVD of matrix A in the following form:

$$A = U\Sigma V^T$$

[2pts]

Hint: Reconstruct matrix A from the SVD to check your answer.

Solution:

The two required products are:

$$AA^T = M = \begin{bmatrix} 18 & 0 \\ 0 & 8 \end{bmatrix}; A^TA = N = \begin{bmatrix} 13 & 5 & 0 \\ 5 & 13 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(a) The eigenvalue equations are:

$$\lambda_M : \begin{vmatrix} 18-\lambda & 0 \\ 0 & 8-\lambda \end{vmatrix} = 0; \lambda_N : \begin{vmatrix} 13-\lambda & 5 & 0 \\ 5 & 13-\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} = 0$$

Solving these equations, we have:

$$\boxed{\lambda_M = 18, 8; \lambda_N = 18, 8, 0}$$

We can now form the scaling matrix Σ as:

$$\boxed{\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 & 0 \\ 0 & 2\sqrt{2} & 0 \end{bmatrix}}$$

(b) The eigenvectors for M are given by the following equations:

$$\begin{bmatrix} 18-\lambda_M & 0 \\ 0 & 8-\lambda_M \end{bmatrix} \cdot v_M = 0$$

Solving these yields the normalized eigenvectors for M as $v_M = (1, 0)$ and $(0, 1)$ correspondingly.

(c) The eigenvectors for N are given by the following equations:

$$\begin{bmatrix} 13-\lambda & 5 & 0 \\ 5 & 13-\lambda & 0 \\ 0 & 0 & -\lambda \end{bmatrix} \cdot v_N = 0;$$

The normalized eigenvectors for N are $v_N = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ and $(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ and $(0, 0, 1)$ correspondingly.

(d) Now, we formulate the SVD as: $\boxed{A = U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 & 0 \\ 0 & 2\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}}$

Verifying by multiplying the matrices, we get the correct answer

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

2 Expectation, Co-variance and Independence [25pts]

Suppose X, Y and Z are three different random variables. Let X obeys Bernouli Distribution. The probability distribution function is

$$p(x) = \begin{cases} 0.5 & x = c \\ 0.5 & x = -c. \end{cases}$$

c is a constant here. Let Y obeys the standard Normal (Gaussian) distribution, which can be written as $Y \sim N(0,1)$. X and Y are independent. Meanwhile, let $Z = XY$.

(a) What is the Expectation and Variance of X?(in terms of c) [4pts]

(b) Show that Z also follows a Normal (Gaussian) distribution. Calculate the Expectation and Variance of Z. [9pts]

(c) How should we choose c such that Y and Z are uncorrelated(which means $Cov(Y, Z) = 0$)? [5pts]

(d) Determine whether the following probability is greater than or equal to 0: (1) $P(Y = 0)$; (2) $P(Z = c)$; (3) $P(Y \in (-1, 0))$; (4) $P(Z \in (2c, 3c))$; (5) $P(Y \in (-1, 0), Z \in (2c, 3c))$; (6) $P(Y \in (-2, -1), Z \in (c, 2c))$.[3pts]

(e) Are Y and Z independent? Make use of the above probabilities to show your conclusion.[4pts]

Solution:

(a) The expectation of X is given by:

$$E[X] = P(X = c) \times c + P(X = -c) \times -c = 0.5c - 0.5c \Rightarrow \boxed{E[X] = 0}$$

The variance is given by:

$$Var(X) = E[X^2] - E[X]^2 = Pr(X = c) \times c^2 + P(X = -c) \times (-c)^2 - 0 \Rightarrow \boxed{Var(X) = c^2}$$

(b) Since $Z = XY$,

$$P(Z < z) = P(XY < z)$$

$$\begin{aligned} P(Z \geq z) &= P(X = c)P(XY \geq z|X = c) + P(X = -c)P(XY \geq z|X = -c) \\ &= \frac{1}{2}P(XY \geq z|X = c) + \frac{1}{2}P(XY \geq z|X = -c) \\ &= \frac{1}{2}P(cY \geq z) + \frac{1}{2}P(-cY \geq z) = \frac{1}{2}P\left(Y \geq \frac{z}{c}\right) + \frac{1}{2}P\left(Y \leq \frac{-z}{c}\right) = P\left(Y \geq \frac{z}{c}\right) \end{aligned}$$

$$P(Z \geq z) = 1 - Q\left(\frac{z}{c}\right)$$

where Q represents the quantile function for $\frac{z}{c}$ in the standard normal distribution. So, from the above equality, we find that $Z \sim N(0, c^2)$. So, we have

$$\boxed{E[Z] = 0; Var(Z) = c^2}$$

(c)

$$Cov(Y,Z) = E[YZ] - E[Y]E[Z]$$

$$\Rightarrow Cov(Y,Z) = P(X=c)E[YZ|X=c] + P(X=-c)E[YZ|X=-c] - 0$$

$$= \frac{1}{2}E[YZ|X=c] + \frac{1}{2}E[YZ|X=-c] = \frac{1}{2}E[cY^2] + \frac{1}{2}E[-cY^2] \Rightarrow \boxed{Cov(Y,Z) = 0 \forall c}$$

As such, we can choose any value for c (except 0, which yields a trivial solution).

(d) (1) Since Y is a normal distribution (which is continuous), $\boxed{P(Y=0)=0}$.

(2) For the same reason, $\boxed{P(Z=c)=0}$.

(3) The probability over an interval is defined for a continuous distribution, so $\boxed{P(Y \in (-1,0)) > 0}$.

(4) For the same reason, $\boxed{P(Z \in (2c,3c)) > 0}$.

(5) Expressing $Z = XY$,

$$P(Y \in (-1,0), Z \in (2c,3c)) = P\left(Y \in (-1,0), \frac{1}{2}Yc \in (2c,3c) + \frac{1}{2}Y(-c) \in (2c,3c)\right)$$

$$= P\left(Y \in (-1,0), \frac{1}{2}Y \in (2,3) - \frac{1}{2}Y \in (2,3)\right)$$

$$\Rightarrow \boxed{P(Y \in (-1,0), Z \in (2c,3c)) = 0}$$

(6) Similarly, this probability is given by:

$$P(Y \in (-2,-1), Z \in (c,2c)) = P\left(Y \in (-2,-1), \frac{1}{2}Yc \in (c,2c) - \frac{1}{2}Yc \in (c,2c)\right)$$

$$\Rightarrow \boxed{P(Y \in (-2,-1), Z \in (c,2c)) > 0}$$

(e) for Y and Z to be independent, we need to have:

$$P(Y \in (-1,0), Z \in (2c,3c)) = P(Y \in (-1,0))P(Z \in (2c,3c))$$

But, both of these probabilities individually are greater than 0 (from the results of part (3) and (4). However, their joint probability is equal to zero. As such, Y and Z are $\boxed{\textit{correlated}}$ random variables.

3 Maximum Likelihood [25 + 10 pts]

3.1 Discrete Example [10 pts]

Suppose we have two types of coins, A and B. The probability of a Type A coin showing heads is θ. The probability of a Type B coin showing heads is 2θ. Here, we have a bunch of coins of either type A or B. Each time we choose one coin and flip it. We do this experiment 10 times and the results are shown in the chart below.

Coin Type	Result
A	Tail
A	Head
A	Tail
B	Head
A	Tail
A	Tail
B	Head
B	Head
B	Head
A	Tail

(a) What is the likelihood of the result given θ? [4pts]

(b) What is the maximum likelihood estimation for θ? [6pts]

Solution:

(a) Since each coin toss is i.i.d., the likelihood of the given result is the product of the individual likelihoods of each coin toss. So, we have:

$$L(result|\theta) = (1-\theta)(\theta)(1-\theta)(2\theta)(1-\theta)(1-\theta)(2\theta)(2\theta)(2\theta)(1-\theta)$$

$$\Rightarrow \boxed{L(result|\theta) = 16\theta^5(1-\theta)^5}$$

(b) The MLE of θ is given by:

$$\frac{dL}{d\theta} = 0$$

The above equation is satisfied for $\boxed{\theta = 0.5}$ which is the maximum likelihood estimation for θ.

3.2 [10 pts]

The C.D.F of independent random variables $X_1, X_2, ..., X_n$ is

$$P(X_i \le x|\alpha, \beta) = \begin{cases} 0, & x < 0 \\ (\frac{x}{\beta})^\alpha, & 0 \le x \le \beta \\ 1, & x > \beta \end{cases}$$

where $\alpha \ge 0$, $\beta \ge 0$. Find the MLEs of α and β.

Solution:
Given the C.D.F of the RVs, we can calculate the P.D.F as:

$$P(x|\alpha, \beta) = \begin{cases} 0, & x < 0 \\ \frac{\alpha \cdot x^{\alpha-1}}{\beta^\alpha}, & 0 \le x \le \beta \\ 0, & x > \beta \end{cases}$$

So, we can define the likelihood function as:

$$f(x_1, x_2 ... x_n|\alpha, \beta) = \prod_{i=1}^{n} \frac{\alpha \cdot x_i^{\alpha-1}}{\beta^\alpha}$$

Seeing that $\beta \ge x_i \forall i$ and that the likelihood decreases with increasing β, we get that:

$$\boxed{\beta_{MLE} = max\{x_i\}}$$

We then obtain the log-likelihood as:

$$l(\alpha, \beta) = n[log(\alpha) - \alpha log(\beta)] + \sum_{i=1}^{n} (\alpha - 1) log(x_i)$$

Solving for the MLE of α,

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - n log(\beta) + \sum_{i=1}^{n} log(x_i) = 0$$

$$\Rightarrow \boxed{\alpha_{MLE} = \frac{n}{n log(max\{x_i\}) - \sum_{i=1}^{n} log(x_i)}}$$

3.3 Poisson distribution [5 pts]

The Poisson distribution is defined as

$$P(x_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}(k = 0, 1, 2, ...).$$

What is the maximum likelihood estimator of λ?

Solution:

The log likelihood of the Poisson distribution is given by:

$$log\left(\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) = -n\lambda - \sum_{i=1}^{n} log(x_i!) + log(\lambda)\sum_{i=1}^{n} x_i$$

Taking the derivative of the R.H.S and setting to 0,

$$-n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \boxed{\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

3.4 Bonus [10 pts]

Given n i.i.d. observations $\{(x_i, y_i)\}_{i=1}^n \in \mathbb{R}^d \times \{-1, 1\}$, we assume

$$\mathbb{P}(y_i = 1|x_i) = h(x_i^T\theta) \text{ and } \mathbb{P}(y_i = -1|x_i) = 1 - h(x_i^T\theta)$$

where $h(x) = \frac{1}{1+\exp(-x)}$ and θ is the model parameter and $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^T$.

Write out the likelihood function $L(\theta)$ given (x_i, y_i). Then formulate the log-likelihood function.

Solution:

The likelihood $L(\theta)$ is given by:

$$L(\theta) = \prod_{i=1}^n h(x_i^T\theta)^{\frac{1+y_i}{2}}(1 - h(x_i^T\theta))^{1-\frac{1+y_i}{2}}$$

Taking the logarithm of the above expression,

$$l(\theta) = \sum_{i=1}^n \left[\frac{1+y_i}{2} log(h(x_i^T\theta)) + \left(1 - \frac{1+y_i}{2}\right) log(1 - h(x_i^T\theta))\right]$$

$$\boxed{l(\theta) = \sum_{i=1}^n \left[\frac{1+y_i}{2} log\left(\frac{1}{1+exp(-x_i^T\theta)}\right) + \left(1 - \frac{1+y_i}{2}\right) log\left(1 - \frac{1}{1+exp(-x_i^T\theta)}\right)\right]}$$

4 Information Theory [25pts + 7pts]

4.1 Marginal Distribution [6pts]

Suppose the joint probability distribution of two binary random variables X and Y are given as follows.

$X \mid Y$	1	2
0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{2}$	0

(a) Show the marginal distribution of X and Y, respectively. [3pts]

(b) Find mutual information for the joint probability distribution in the previous question [3pts]

Solution:

(a) The marginal distribution of X is given by:

x	$P(X = x)$
0	$\frac{1}{2}$
1	$\frac{1}{2}$

and the marginal distribution of Y is given by:

y	$P(Y = y)$
1	$\frac{3}{4}$
2	$\frac{1}{4}$

(b) The mutual information of the joint probability distribution is given by:

$$I(X, Y) = \sum_x \sum_y p_{X,Y}(x, y) log \left(\frac{p_{X,Y}(x, y)}{p_X(x) p_Y(y)} \right)$$

$$\Rightarrow I(X, Y) = \frac{1}{4} log \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{3}{4}} \right) + \frac{1}{4} log \left(\frac{\frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{4}} \right) + \frac{1}{2} log \left(\frac{\frac{1}{2}}{\frac{1}{2} \cdot \frac{3}{4}} \right) + 0$$

$$\Rightarrow \boxed{I(X, Y) = 0.311}$$

4.2 Mutual Information and Entropy [19pts]

Given a dataset as below.

Player	*Experience*	*NumUtilities*	*BuysBoardwalk?*	*Hunger*	*Outcome*
1	*novice*	2	*no*	*low*	*lose*
2	*intermediate*	0	*no*	*high*	*lose*
3	*novice*	1	*no*	*low*	*win*
4	*expert*	0	*no*	*medium*	*win*
5	*intermediate*	0	*yes*	*high*	*win*
6	*expert*	0	*yes*	*high*	*lose*
7	*intermediate*	2	*yes*	*low*	*win*
8	*intermediate*	1	*no*	*medium*	*win*
9	*expert*	1	*no*	*low*	*lose*
10	*novice*	0	*no*	*medium*	*lose*
11	*novice*	2	*yes*	*low*	*win*
12	*intermediate*	1	*no*	*medium*	*lose*
13	*intermediate*	0	*yes*	*high*	*win*
14	*novice*	0	*yes*	*high*	*lose*

You are analyzing data from your last few Monopoly games in hopes of becoming a world champion. We want to determine what makes a player win or lose. Each input has four features (x_1, x_2, x_3, x_4): Experience, NumUtilities, BuysBoardwalk, Hunger. The outcome (win vs lose) is represented as Y.

(a) Find entropy $H(Y)$. [3pts]

(b) Find conditional entropy $H(Y|x_1)$, $H(Y|x_4)$, respectively. [8pts]

(c) Find mutual information $I(x_1, Y)$ and $I(x_4, Y)$ and determine which one (x_1 or x_4) is more informative. [4pts]

(d) Find joint entropy $H(Y, x_3)$. [4pts]

Solution:

(a) The entropy $H(Y)$ is given by:

$$H(Y) = -[P(win)log(P(win)) + P(lose)log(P(lose))]$$

$$\Rightarrow H(Y) = -\left[\frac{7}{14}log\left(\frac{7}{14}\right) + \frac{7}{14}log\left(\frac{7}{14}\right)\right]$$

$$\Rightarrow \boxed{H(Y) = 1}$$

(b) The conditional entropy $H(Y|x_1)$ is given by:

$$-\sum_{x\in x_1;y\in Y} P(x,y)log\left(\frac{P(x,y)}{P(x)}\right)$$

$$= -[P(n,l)log\left(\frac{P(n,l)}{P(n)}\right) + P(i,l)log\left(\frac{P(i,l)}{P(i)}\right) + P(e,l)log\left(\frac{P(e,l)}{P(e)}\right) + P(n,w)log\left(\frac{P(n,w)}{P(n)}\right)$$
$$+P(i,w)log\left(\frac{P(i,w)}{P(i)}\right) + P(e,w)log\left(\frac{P(e,w)}{P(e)}\right)]$$
$$= -\left[3log\left(\frac{3}{5}\right) + 2log\left(\frac{2}{6}\right) + 2log\left(\frac{2}{3}\right) + 2log\left(\frac{2}{5}\right) + 4log\left(\frac{4}{6}\right) + 1log\left(\frac{1}{3}\right)\right] \cdot \frac{1}{14}$$
$$\Rightarrow \boxed{H(Y|x_1) = 0.937}$$

Similarly, the conditional entropy $H(Y|x_4)$ is given by:

$$-\sum_{x \in x_4; y \in Y} P(x,y)log\left(\frac{P(x,y)}{P(x)}\right)$$

$$= -[P(low,l)log\left(\frac{P(low,l)}{P(low)}\right) + P(m,l)log\left(\frac{P(m,l)}{P(m)}\right) + P(h,l)log\left(\frac{P(h,l)}{P(h)}\right) + P(low,w)log\left(\frac{P(low,w)}{P(low)}\right)$$
$$+P(m,w)log\left(\frac{P(m,w)}{P(m)}\right) + P(h,w)log\left(\frac{P(h,w)}{P(h)}\right)]$$
$$= -\left[2log\left(\frac{2}{5}\right) + 2log\left(\frac{2}{4}\right) + 3log\left(\frac{3}{5}\right) + 3log\left(\frac{3}{5}\right) + 2log\left(\frac{2}{4}\right) + 2log\left(\frac{2}{5}\right)\right] \cdot \frac{1}{14}$$
$$\Rightarrow \boxed{H(Y|x_4) = 0.979}$$

(c) The mutual information values are equivalently expressed as $I(Y, x_1)$ and $I(Y, x_4)$.
Using the property
$$I(Y, X) = H(Y) - H(Y|X)$$

So, we have:

$$\boxed{I(x_1, Y) = 1 - 0.937 = 0.063}; \boxed{I(x_4, Y) = 1 - 0.979 = 0.021}$$

Which shows that x_1 is more informative.

(d) The joint entropy $H(Y, x_3)$ is given by:

$$-\sum_{x \in x_3; y \in Y} P(x,y)log(P(x,y))$$

$$= -\left[P(y,w)log(P(y,w)) + P(n,w)log(P(n,w)) + P(y,l)log(P(y,l)) + P(n,l)log(P(n,l))\right]$$
$$= -\left[\frac{4}{14}log\left(\frac{4}{14}\right) + \frac{3}{14}log\left(\frac{3}{14}\right) + \frac{2}{14}log\left(\frac{2}{14}\right) + \frac{5}{14}log\left(\frac{5}{14}\right)\right]$$
$$\Rightarrow \boxed{H(Y, x_3) = 1.924}$$

4.3 Bonus Question [7pts]

(a) Suppose X and Y are independent. Show that $H(X|Y) = H(X)$. [2pts]

(b) Suppose X and Y are independent. Show that $H(X,Y) = H(X) + H(Y)$. [2pts]

(c) Prove that the mutual information is symmetric, i.e., $I(X,Y) = I(Y,X)$ and $x_i \in X, y_i \in Y$ [3pts]

Solution:

(a) The conditional entropy of X on Y is given by:

$$H(X|Y) = -\sum_{i,j} P(X = x_i, Y = y_j) log\left(\frac{P(X = x_i, Y = y_i)}{P(Y = y_j)}\right)$$

$$= -\sum_{i,j} P(X = x_i)P(Y = y_i) log\left(\frac{P(X = x_i)P(Y = y_j)}{P(Y = y_j)}\right)$$

$$= -\sum_{i} [P(x_i)log(P(x_i))] \cdot \sum_{j} P(Y = y_j) = -\sum_{i} P(x_i)log(P(x_i))$$

which is the the entropy of the R.V. X.

(b) The joint entropy of X and Y is given by:

$$H(X,Y) = -\sum_{i,j} P(X = x_i, Y = y_j) log(P(X = x_i, Y = y_j))$$

$$= -\sum_{i,j} P(X = x_i)P(Y = y_j) log(P(X = x_i)P(Y = y_j))$$

$$= -\sum_{i,j} P(X = x_i)P(Y = y_j)[log(P(X = x_i)) + log(P(Y = y_j))]$$

$$= -\sum_{i} P(X = x_i)log(P(X = x_i)) \sum_{j} P(Y = y_j) - \sum_{i} P(Y = y_j)log(P(Y = y_j)) \sum_{i} P(X = x_i)$$

$$= -\sum_{i} P(X = x_i)log(P(X = x_i)) - \sum_{i} P(Y = y_j)log(P(Y = y_j)) = H(X) + H(Y)$$

(c) The mutual information $I(X,Y)$ is given by:

$$I(X,Y) = \sum_{i}\sum_{j} P(X = x_i, Y = y_j) log\left(\frac{P(X = x_i, Y = y_j)}{P(X = x_i)P(Y = y_j)}\right)$$

Interchanging the summation order in the above expression,

$$I(X,Y) = \sum_{j}\sum_{i} P(Y = y_j, X = x_i) log\left(\frac{P(Y = y_j, X = x_i)}{P(Y = y_j)P(X = x_i)}\right) = I(Y,X)$$

5 Bonus for All [10 pts]

Due to the recent social distancing requirement, Wal-Mart is re-evaluating their delivery policies. In order to properly update their policy, Wal-Mart is analyzing data from previous records. Delivery time can be classified as early, on time or late. Delivery distance can be classified as within 5 miles, between 5 and 10 miles and over 10 miles. From the previous records, 15% of deliveries arrive early, and 55% arrive on time. 70% of orders are within 5 miles and 25% of orders are between 5 and 10 miles. The probability for arriving on time if delivery distance is over 10 miles is 0. The probability of a shipment arriving on time and having a delivery distance between 5 and 10 miles is 10%. The probability for arriving early if delivery distance is within 5 miles is 20%.

(a) What is the probability that the delivery will arrive on time if the distance is between 5 and 10 miles? [2 pts]

(b) What is the probability that the delivery will arrive on time if the distance is within 5 miles? [4 pts]

(c) What is the probability that the delivery will arrive late if the distance is within 5 miles? [4 pts]

Solution:

The probabilities mentioned can be expressed as:

$$P(e) = 0.15$$
$$P(o) = 0.55$$
$$P(l) = 1 - P(e) - P(o) = 0.3$$
$$P(d < 5) = 0.7$$
$$P(5 < d < 10) = 0.25$$
$$P(d > 10) = 1 - P(d < 5) - P(5 < d < 10) = 0.05$$

The conditional and joint probabilities are given by:

$$P(o|(d > 10)) = 0$$
$$P(o \cap (5 < d < 10)) = 0.1$$
$$P(e|d < 5) = 0.2$$

(a) The required probability is expressed as:

$$P(o|5 < d < 10) = \frac{P(o \cap (5 < d < 10))}{P(5 < d < 10)}$$
$$\Rightarrow \boxed{P(o|5 < d < 10) = 0.4}$$

(b) From the above result, we find that:

$$P(o) = P(o|d < 5) \cdot P(d < 5) + P(o|(5 < d < 10) \cdot P(5 < d < 10) + P(o|d > 10) \cdot P(d > 10)$$
$$0.55 = 0.7 \cdot P(o|d < 5) + 0.4 \cdot 0.25 + 0$$
$$\Rightarrow \boxed{P(o|d < 5) = 0.643}$$

(c) The required probability is $P(l|d<5)$ which is given by the equation:

$$P(d<5) = P(d<5|e)P(e) + P(d<5|o)P(o) + P(d<5|l)P(l)$$

So,

$$P(d<5) = P(e|d<5)\frac{P(d<5)}{P(e)}P(e) + P(o|d<5)\frac{P(d<5)}{P(o)}P(o) + P(l|d<5)\frac{P(d<5)}{P(l)}P(l)$$

$$\Rightarrow 0.7 = 0.2(0.7) + 0.643(0.7) + P(l|d<5)0.7$$

$$\Rightarrow \boxed{P(l|d<5) = 0.157}$$