Advanced Machine Learning course
Home assignment 2

Yaniv Tal, 031431166

March 16, 2022

1 Question 1

1.1 part a.

Mutual information is defined as the relative entropy between the RVs joint distribution p(x, y) and the product distribution p(x)p(y):

$$I(X;Y) = \sum_{x\in X, y\in Y} p_{(X,Y)}(x,y)\log\frac{p_{(X,Y)}(x,y)}{p_X(x)p_Y(y)} =$$

$$\sum_{x\in X, y\in Y} p_{(X,Y)}(x,y)\log\frac{p_{(X,Y)}(x,y)}{p_X(x)} - \sum_{x\in X, y\in Y} p_{(X,Y)}(x,y)\log p_Y(y) =$$

$$\sum_{x\in X, y\in Y} p_X(x)p_{Y|X=x}(y)\log p_{Y|X=x}(y) - \sum_{x\in X, y\in Y} p_{(X,Y)}(x,y)\log p_Y(y) =$$

$$\sum_{x\in X} p_X(x)\left(\sum_{y\in Y} p_{Y|X=x}(y)\log p_{Y|X=x}(y)\right) - \sum_{y\in Y}\left(\sum_{x\in X} p_{(X,Y)}(x,y)\right)\log p_Y(y) =$$

$$-\sum_{x\in X} p_X(x)H(Y \mid X=x) - \sum_{y\in Y} p_Y(y)\log p_Y(y) =$$

$$-H(Y \mid X) + H(Y) =$$

$$H(Y) - H(Y \mid X)$$

Since $I(X;Y)$ is symmetrical for x, y, we can do the exact same process with y instead of x and get:

$$I(X;Y) = H(X) - H(X \mid Y).$$

1.2 part b.

i. Required:

$$I(X;Y \mid Z) < I(X;Y)$$

Writing as sums of entropy:

$$H(X \mid Z) - H(X \mid Y, Z) < H(X) - H(X \mid Y)$$

An example distribution that fulfills this requirement can be seen in the table below:

Z	X	Y	Pr(x, y, z)
0	0	0	0.1
0	0	1	0
0	1	0	0.2
0	1	1	0
1	0	0	0
1	0	1	0.4
1	1	0	0
1	1	1	0.3

Table 1: $I(X;Y \mid Z) < I(X;Y)$

Calculating the above entropy elements one by one:

Entropy of X:

$$H(X) = \sum_x p(x)log(p(x)) = \big(-0.5log(0.5)\big) + \big(-0.5log(0.5)\big) = 1$$

Entropy of $(X \mid Y)$ - expected to be the same since Y adds no information on X:

$$H(X \mid Y) = \sum_Y p(y)H(X \mid y) = -\sum_y \sum_x p(x,y)log\big(p(x \mid y)\big) =$$

$$\big(-0.1)log(1/3)\big) + \big(-0.2)log(2/3)\big) + \big(-0.4)log(4/7)\big) + \big(-0.3)log(3/7)\big) = 0.965$$

Entropy of $(X \mid Z)$ - expected to be the same since Z adds no information on X:

$$H(X \mid Z) = \sum_z p(z)H(X \mid z) = -\sum_y \sum_x p(x,z)log\big(p(x \mid z)\big) =$$

$$\big(-0.1)log(1/3)\big) + \big(-0.2)log(2/3)\big) + \big(-0.4)log(4/7)\big) + \big(-0.3)log(3/7)\big) = 0.965$$

Entropy of $(X \mid Y, Z)$ - Expected to be zero since Y, Z define the value of X uniquely:

$$H\big(X \mid Y, Z\big) = \sum_y \sum_z p(y,z)H(X \mid y,z) =$$

$$-\sum_x \sum_y \sum_z p(x,y,z)log\big(p(x \mid y,z)\big) =$$

$$\big(-0.1)log(1/3)\big) + \big(-0.2)log(2/3)\big) + \big(-0.4)log(4/7)\big) + \big(-0.3)log(3/7)\big) = 0.965$$

$$I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z) = 0$$
$$I(X;Y) = H(X) - H(X \mid Y) = 1 - 0.965 = 0.035$$

And as required:

$$I(X;Y \mid Z) < I(X;Y)$$

ii. Now on the opposite direction - Required:

$$I(X;Y \mid Z) > I(X;Y)$$

Writing as sums of entropy:

$$H(X \mid Z) - H(X \mid Y,Z) > H(X) - H(X \mid Y)$$

An example distribution that fulfills this requirement can be seen in the table below:

Z	X	Y	Pr(x, y, z)
0	0	0	0
0	0	1	0.25
0	1	0	0.25
0	1	1	0
1	0	0	0.25
1	0	1	0
1	1	0	0
1	1	1	0.25

Table 2: $I(X;Y \mid Z) > I(X;Y)$

Calculating the above entropy elements one by one:

Entropy of X:

$$H(X) = \sum_x p(x)log(p(x)) = 2 \cdot \big(- 0.5log(0.5)\big) = 1$$

Entropy of $(X \mid Y)$ - expected to be the same since Y adds no information on X:

$$H(X \mid Y) = \sum_Y p(y)H(X \mid y) = -\sum_y \sum_x p(x,y)log\big(p(x \mid y)\big) = 4 \cdot \big(- 0.25log(0.5)\big) = 1$$

Entropy of $(X \mid Z)$ - expected to be the same since Z adds no information on X:

$$H(X \mid Z) = \sum_z p(z)H(X \mid z) = -\sum_y \sum_x p(x,z)log\big(p(x \mid z)\big) = 4 \cdot \big(- 0.25log(0.5)\big) = 1$$

Entropy of $(X \mid Y,Z)$ - Expected to be zero since Y,Z define the value of X uniquely:

$$H\big(X \mid Y,Z\big) = \sum_y \sum_z p(y,z)H(X \mid y,z) =$$

$$-\sum_x \sum_y \sum_z p(x,y,z)log\big(p(x \mid y,z)\big) = 0$$

$$I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y,Z) = 1 - 0 = 1$$
$$I(X;Y) = H(X) - H(X \mid Y) = 1 - 1 = 0$$

And as required:

$$I(X;Y \mid Z) > I(X;Y)$$

2 Question 2

Let X, Y, Z three random variables who form a Markov chain $X \to Y \to Z$. Based on Bayes:

$$p(x, z \mid y) = \frac{p(y \mid x, z)p(x, z)}{p(y)} = \frac{p(x, y, z)}{p(y)}$$

From the Markovian property of the chain, we get that:

$$p(x, y, z) = p(x)p(y \mid x)p(z \mid y)$$

putting it together and using Bayes to get $p(x \mid y$:

$$p(x, z \mid y) = \frac{p(x)p(y \mid x)p(z \mid y)}{p(y)} = p(x \mid y)p(z \mid y)$$

As required

3 Question 3

$Pr(x_1, x_2)$	x_1	x_2
y_1	1/4	0
y_2	1/4	1/2

Table 3:

a. $H(X), H(Y)$

$$H(X) = \sum_x -p(x)log\big(p(x)\big) = -2 \cdot 0.5log(0.5) = 1$$

$$H(Y) = \sum_x -p(x)log\big(p(x)\big) = -0.25log(0.25) - 0.75log(0.75) = 0.811$$

b. $H(X \mid Y), H(Y \mid X)$

$$H(X \mid Y) = \sum_Y p(y)H(X \mid Y) = \sum_Y \sum_X p(x,y)log\big(p(x \mid y)\big) =$$

$$-0.25log(1) - 0 - 0.25log(0.333) - 0.5log(0.666) = 0.688$$

$$H(Y \mid X) = \sum_X p(x)H(Y \mid X) = \sum_X \sum_Y p(x,y)log\big(p(y \mid x)\big) =$$

$$-0.25log(0.5) - 0.25log(0.5) - 0 - 0.5log(1) = 0.5$$

c. $H(X, Y)$

$$H(X,Y) = -0.25log(0.25) - 0.25log(0.25) - 0 - 0.5log(0.5) = 1.5$$

d. $H(X) - H(Y)$

$$H(X) - H(Y) = 1 - 0.811 = 0.189$$

e. $I(X;Y)$

$$I(X;Y) = H(X) - H\big(X \mid Y\big) = 1 - 0.688 = 0.312$$

f. Venn diagram:

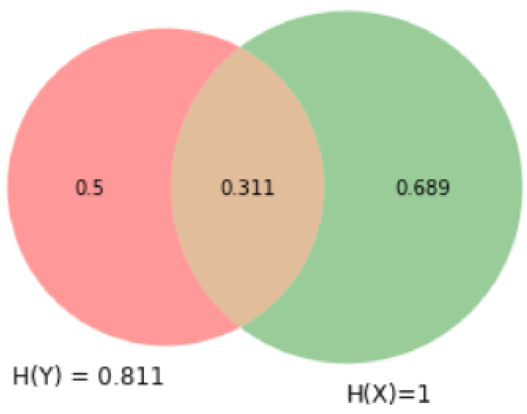

Figure 1: Venn diagram

4 Question 4

Let

$$p = (p_1, p_2, ..., p_m), \sum p_i = 1$$

Also, let:

$$q = (q_1, q_2, ..., q_{m-1}), \sum p_i = 1$$

such that:

$$q_1 = p_1, q_2 = p_2 \; ..... \; q_{m-2} = p_{m-2}, q_{m-1} = (p_{m-1} + p_m)$$

Also let:

$$v = \Big(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\Big)$$

We need to show that: $H(q) + (p_{m-1} + p_m)H(v) = H(p)$.

First, lets look at $H(v)$:

$$H(v) = -\Bigg[\frac{p_{m-1}}{p_{m-1} + p_m} log\Big(\frac{p_{m-1}}{p_{m-1} + p_m}\Big) + \frac{p_m}{p_{m-1} + p_m} log\Big(\frac{p_m}{p_{m-1} + p_m}\Big)\Bigg] =$$

$$-\Bigg[\frac{p_{m-1}}{p_{m-1} + p_m}\Big(log(p_{m-1}) - log(p_{m-1+p_m})\Big) + \frac{p_m}{p_{m-1} + p_m}\Big(log(p_m) - log(p_{m-1} + p_m)\Big)\Bigg] =$$

$$log(p_{m-1} + p_m) - \frac{p_{m-1}log(p_{m-1}) + p_m log(p_m)}{p_{m-1} + p_m}$$

Now lets look at H(q):

$$H(q) = \sum_{i=1}^{m-2} \big(- p_i log(p_i)\big) - (p_{m-1} + p_m)log(p_{m-1} + p_m)$$

Putting it all together:

$$H(q) + (p_{m-1} + p_m)H(v) =$$

$$\sum_{i=1}^{m-2} \big(-p_i log(p_i)\big) - (p_{m-1}+p_m)log(p_{m-1}+p_m) + (p_{m-1}+p_m)log(p_{m-1}+p_m) - \Big(p_{m-1}log(p_{m-1}) + p_m log(p_m)\Big) =$$

$$\sum_{i=1}^{m-2} \big(- p_i log(p_i)\big) - p_{m-1}log(p_{m-1}) - p_m log(p_m) =$$

$$\sum_{i=1}^{m} \big(- p_i log(p_i)\big) = H(p)$$

As required

5 Question 5

Generally, we are looking for distributions $p(x) \neq q(x)$ that will fulfill the following:

$$\sum_x -p(x)log\Big(\frac{q(x)}{p(x)}\Big) = \sum_x -q(x)log\Big(\frac{p(x)}{q(x)}\Big)$$

I'll choose a Bernully dist. for simplicity, than:

$$-p(x)log\Big(\frac{q(x)}{p(x)}\Big) - (1-q(x))log\Big(\frac{1-q(x)}{1-p(x)}\Big) = -q(x)log\Big(\frac{p(x)}{q(x)}\Big) - (1-p(x))log\Big(\frac{1-p(x)}{1-q(x)}\Big)$$

The solutions are either $p(x) = q(x)$, or $p(x) = 1 - q(x)$.

So the example is: $p \sim Benrully(0.3)$**,** $q \sim Benrully(0.7)$
It is obvious that the following is true:

$$D(p\,||\,q) = -0.3log\Big(\frac{0.7}{0.3}\Big) - 0.7log\Big(\frac{0.3}{0.7}\Big) = -0.7log\Big(\frac{0.3}{0.7}\Big) - 0.3log\Big(\frac{0.7}{0.3}\Big) = D(q\,||\,p)$$

6 Question 6

By definition, the Taylor expansion for f around c is:

$$f(x) = f(c) + f'(c)(x-c) + f''(c)\frac{(x-c)^2}{2} + ...$$

Now expanding the KL divergence around q:

$$D(p \,||\, q) = D(p \,||\, q)_{p=q} + D'(p \,||\, q)_{p=q} + D''(p \,||\, q)_{p=q} + ...$$

Calculate each element, use natural log for convenience:
First derivative:

$$D(p \,||\, q)_{p=q} = plog\frac{p}{q} = 0log(0) = 0$$

$$\frac{\partial}{\partial p} D(p \,||\, q)_{p=q} = \frac{\partial}{\partial p} \sum plog(\frac{p}{q}) =$$

$$\frac{\partial}{\partial p} \sum \Big(plog(p) - plog(q)\Big)_{p=q} = \Big(log(p) + p\frac{1}{p} - log(q)\Big)_{p=q} =$$

$$log(\frac{p}{q})_{p=q} + 1 = 1$$

Second derivative:

$$\frac{\partial^2}{\partial p^2} D(p \,||\, q)_{p=q} = \frac{\partial}{\partial p} \sum \Big(log(\frac{p}{q}) + 1\Big)_{p=q} = \sum \Big(\frac{1}{p}\Big)_{p=q} = \sum \frac{1}{q}$$

Put it together:

$$D(p \,||\, q) = 0 + \sum 1(p-q) + \sum \frac{1}{q} \frac{(p-q)^2}{2} + ... =$$

We know that:

$$\chi^2 = \sum \frac{(p-q)^2}{q}$$

So the taylor sum becomes:

$$0 + 0 + \frac{1}{2}\chi^2 + ...$$

As required

7 Question 7

Need to find:

$$p* = ArgminD(p \mid\mid q)$$

$$\textbf{s.t.}$$

$$\forall j : \sum_x p(x)f_j(x) = \alpha_j$$

Using Lagrange:

$$\mathcal{L}(p, \lambda) = \sum_x plog\frac{p}{q} + \sum_j \Big(\lambda_j(\sum_x p(x)f_j(x) - \alpha_j)\Big)$$

Find $\frac{\partial}{\partial p}\mathcal{L}$ and set to zero:

$$\frac{\partial}{\partial p}\mathcal{L}(p, \lambda) = log(\frac{p}{q}) + 1 + \sum_j \lambda_j f_j(x) =$$

$$log(p) - log(q) + 1 + \sum_j \lambda_j f_j(x) = 0$$

Extracting p:

$$p* = \frac{q(x)}{e^{1+\sum_j \lambda_j f_j(x)}}$$

Find $\frac{\partial}{\partial \lambda_j}\mathcal{L}$ and set to zero for all j's:

$$\frac{\partial}{\partial \lambda_j}\mathcal{L} = \sum_x p(x)f_j(x) - \alpha_j = 0$$

$$\forall j : \alpha_j = \sum_x p(x)f_j(x)$$