# ML-2

## Homework Assignment - 2

Shantanu Chandra
(12048461)

**Q1.**

1. Mutual Inform. of $I(X:Y)$ is defined as:

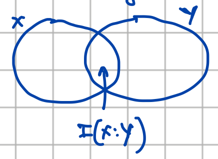
$$I(X:Y) = KL(p(X,Y) \,||\, p(X)\cdot p(Y))$$

Using definition of KL:

(discrete) $$I(X:Y) = \sum_x \sum_y p(x,y) \ln \frac{p(x,y)}{p(x)\,p(y)}$$

(continuous) $$I(X:Y) = \int_x \int_y p(x,y) \ln \frac{p(x,y)}{p(x)\,p(y)} \,dx\,dy$$

This intuitively says how much one random variable tells us about the other, OR in our case how much we reduce uncertainty of X if we knew Y. This implies that if X & Y are dependent, how much information is shared between them and can be inferred in the form of a Venn Diag.:




Conditional mutual information $I(X:Y|Z)$ is defined as:

$$I(X:Y|Z) = KL(p(x,y,z) \,||\, p(x|z)\,p(y|z)\,p(z))$$

Using definition of KL:

(discrete) $$I(X:Y|Z) = \sum_z p(z) \sum_x \sum_y p(x,y|z) \ln \frac{p(x,y|z)}{p(x|z)\,p(y|z)}$$

(continuous) $$I(X:Y|Z) = \int_z p(z) \int_x \int_y p(x,y|z) \ln \frac{p(x,y|z)}{p(x|z)\,p(y|z)} \,dx\,dy$$

This intuitively measures the mutual inform. between 2 distributions/random variables given a third. Hence, if both the distributions are independent of the third, $CMI(X:Y|Z) = MI(X:Y)$

// Ans.

2

| $x$ | $y$ | $z$ | $p(x,y,z)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.192 |
| 0 | 0 | 1 | 0.144 |
| 0 | 1 | 0 | 0.048 |
| 0 | 1 | 1 | 0.216 |
| 1 | 0 | 0 | 0.192 |
| 1 | 0 | 1 | 0.064 |
| 1 | 1 | 0 | 0.048 |
| 1 | 1 | 1 | 0.096 |

| $x$ | $p(x)$ |
|---|---|
| 0 | 0.6 |
| 1 | 0.4 |

| $y$ | $p(y)$ |
|---|---|
| 0 | 0.592 |
| 1 | 0.408 |

| $z$ | $p(z)$ |
|---|---|
| 0 | 0.48 |
| 1 | 0.52 |

| $x$ | $y$ | $p(x,y)$ |
|---|---|---|
| 0 | 0 | 0.336 |
| 0 | 1 | 0.264 |
| 1 | 0 | 0.256 |
| 1 | 1 | 0.144 |

where we have used,

$$p(x) = p(x, y=0) + p(x, y=1)$$
$$p(y) = p(y, z=0) + p(y, x=1)$$
$$p(x,y) = p(x,y,z=0) + p(x,y,z=1)$$

Now,

$$I(X:Y) = KL(p(x,y) \,||\, p(x)\cdot p(y)) = \sum_x \sum_y p(x,y) \ln \frac{p(x,y)}{p(x)\,p(y)}$$

an example of this calculation for $x=0$, $y=0$ :

$$I(x=0 : Y=0) = 0.336 \times \ln \frac{(0.336)}{(0.6)(0.592)} = 0.336 \times (-0.0555)$$
$$= -0.01867$$

Ans.

Doing this for all $x, y$ we get :

| $x$ | $y$ | $I(x,y)$ |
|---|---|---|
| 0 | 0 | -0.0187 |
| 0 | 1 | 0.0199 |
| 1 | 0 | 0.02 |
| 1 | 1 | -0.0180 |

Ans.

Thus, $I(X:Y) = \sum I(x,y)$
$= 0.0032 > 0$

Proved

$I(X:Y) > 0$ implies that there is some shared information between the distributions $p(x)$ & $p(y)$. This in turn means that the random variables X & Y are dependent

More formally we can also say : $I(X:Y) > 0$
or, $H(X) > H(X|Y)$

This means that the entropy (or uncertainty) of a variable reduces by knowing the $2^{nd}$ variable.

Ans.

3

| x | y | z | p(x\|z) | p(y\|z) | p(x,y\|z) |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.5 | 0.8 | 0.4 |
| 0 | 0 | 1 | 0.692 | 0.4 | 0.277 |
| 0 | 1 | 0 | 0.5 | 0.2 | 0.1 |
| 0 | 1 | 1 | 0.692 | 0.6 | 0.415 |
| 1 | 0 | 0 | 0.5 | 0.8 | 0.4 |
| 1 | 0 | 1 | 0.308 | 0.4 | 0.123 |
| 1 | 1 | 0 | 0.5 | 0.2 | 0.1 |
| 1 | 1 | 1 | 0.308 | 0.6 | 0.185 |

where we use,

$$p(x|z) = \sum_y p(x,y|z) = p(x, y=0|z) + p(x, y=1|z)$$

$$p(y|z) = \sum_x p(x,y|z) = p(y, x=0|z) + p(y, x=1|z)$$

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

Now, $I(X:Y|Z) = \sum_z p(z) \sum_x \sum_y p(x,y|z) \ln \frac{p(x,y|z)}{p(x|z)\,p(y|z)}$

an example of this calculation for $x=0, y=0, z=0$ :

$$I(X=0:Y=0\,|\,Z=0) = 0.48 \times 0.4 \ln \frac{0.4}{0.5 \times 0.8}$$
$$= 0.48 \times 0.4 \ln 1$$
$$= 0.48 \times 0$$
$$= \underline{\underline{0}}$$

| x | y | z | I(X:Y\|Z) |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 |

// Ans.

Conditional mutual infor. = 0 implies that Y does not provide any additional inform. about X that can be obtained from Z already.

More formally, $I(X:Y|Z) = 0$
$\Rightarrow H(X|Z) = H(X|Y,Z)$

Thus, <u>X & Y are independent</u> given Z

<u>NOTE</u>: X & Y are dependent in absence of Z as seen in previous question.

// Ans.

4

| x | y | z | p(z\|x) | p(y\|z) | p(z,y\|x) | p(x) | p(z\|x)·p(y\|z)·p(x) |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.4 | 0.8 | 0.32 | 0.6 | 0.192 |
| 0 | 0 | 1 | 0.6 | 0.4 | 0.24 | 0.6 | 0.144 |
| 0 | 1 | 0 | 0.4 | 0.2 | 0.8 | 0.6 | 0.048 |
| 0 | 1 | 1 | 0.6 | 0.6 | 0.36 | 0.6 | 0.216 |
| 1 | 0 | 0 | 0.6 | 0.8 | 0.48 | 0.4 | 0.192 |
| 1 | 0 | 1 | 0.4 | 0.4 | 0.16 | 0.4 | 0.064 |
| 1 | 1 | 0 | 0.6 | 0.2 | 0.12 | 0.4 | 0.048 |
| 1 | 1 | 1 | 0.4 | 0.6 | 0.24 | 0.4 | 0.096 |

where we use,

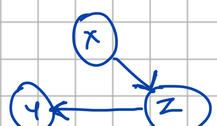
$$p(z|x) = \sum_y p(z,y|x) = p(z, y=0|x) + p(z, y=1|x)$$

$$p(y|z) = \sum_x p(x,y|z) = p(y, x=0|z) + p(y, x=1|z)$$

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

From the table above we see that
$p(x,y,z) = p(z|x) \cdot p(y|z) \cdot p(x)$ — (1)
// Proved

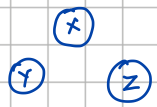
The graph for (1) looks like :




// Ans.

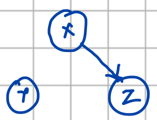
## Q2.

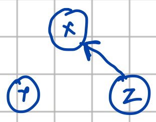
cluster 1 :




Here, $X \perp\!\!\!\perp Y$   $X \perp\!\!\!\perp Y \mid Z$
$X \perp\!\!\!\perp Z$   $X \perp\!\!\!\perp Z \mid Y$
$Y \perp\!\!\!\perp Z$   $Y \perp\!\!\!\perp Z \mid X$

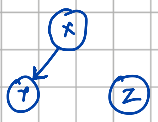
cluster 2 :

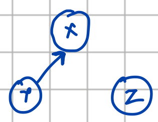





Here,
$X \perp\!\!\!\perp Y$   $X \perp\!\!\!\perp Y \mid Z$
$Y \perp\!\!\!\perp Z$   $Y \perp\!\!\!\perp Z \mid X$
$X \not\perp\!\!\!\perp Z$   $X \not\perp\!\!\!\perp Z \mid Y$

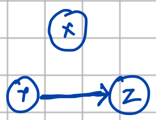
cluster 3 :

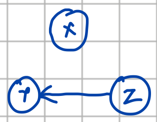





Here, $X \not\perp\!\!\!\perp Y$   $X \not\perp\!\!\!\perp Y \mid Z$
$X \perp\!\!\!\perp Z$   $X \perp\!\!\!\perp Z \mid Y$
$Y \perp\!\!\!\perp Z$   $Y \perp\!\!\!\perp Z \mid X$

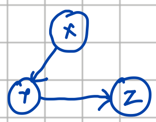
cluster 4 :

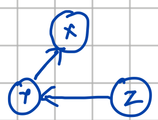




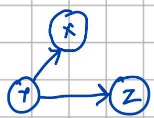

Here, $X \perp\!\!\!\perp Y$   $X \perp\!\!\!\perp Y \mid Z$
$X \perp\!\!\!\perp Z$   $X \perp\!\!\!\perp Z \mid Y$
$Y \not\perp\!\!\!\perp Z$   $Y \not\perp\!\!\!\perp Z \mid X$

cluster 5 :





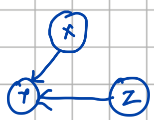

Here, $X \not\perp\!\!\!\perp Y$   $X \not\perp\!\!\!\perp Y \mid Z$
$Y \not\perp\!\!\!\perp Z$   $Y \not\perp\!\!\!\perp Z \mid X$
$X \not\perp\!\!\!\perp Z$   $X \perp\!\!\!\perp Z \mid Y$




Here, $X \not\perp\!\!\!\perp Z$ and $X \perp\!\!\!\perp Z \mid Y$ comes from common cause relation. If common cause (ie, Y) is known then all its effects (ie, X & Z) are known through it without any information from any other. Also, if the common cause is unknown, since all its effects are linked to a common cause, they are all related & thus not independent.

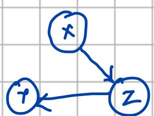
NOTE : similar reasoning follows for the other clusters too.

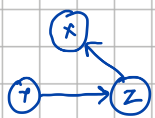
cluster 6 :



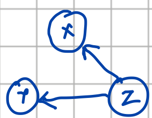

Here, $X \not\perp\!\!\!\perp Y$   $X \not\perp\!\!\!\perp Y \mid Z$
$Y \not\perp\!\!\!\perp Z$   $Y \not\perp\!\!\!\perp Z \mid X$
$X \perp\!\!\!\perp Z$   $X \not\perp\!\!\!\perp Z \mid Y$

cluster 7 :

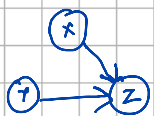





Here, $X \not\perp\!\!\!\perp Y$   $X \perp\!\!\!\perp Y \mid Z$
$X \not\perp\!\!\!\perp Z$   $X \not\perp\!\!\!\perp Z \mid Y$
$Y \not\perp\!\!\!\perp Z$   $Y \not\perp\!\!\!\perp Z \mid X$

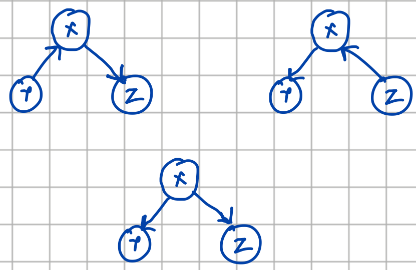



cluster 8 :




$X \perp\!\!\!\perp Y$   $X \not\perp\!\!\!\perp Y \mid Z$
$X \not\perp\!\!\!\perp Z$   $X \not\perp\!\!\!\perp Z \mid Y$
$Y \not\perp\!\!\!\perp Z$   $Y \not\perp\!\!\!\perp Z \mid X$

cluster 9 :



Here,

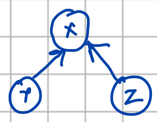
$X \not\perp\!\!\!\perp Y$ $\quad$ $X \not\perp\!\!\!\perp Y \mid Z$

$X \not\perp\!\!\!\perp Z$ $\quad$ $X \not\perp\!\!\!\perp Z \mid Y$

$Y \not\perp\!\!\!\perp Z$ $\quad$ $Y \perp\!\!\!\perp Z \mid X$

---

cluster 10 :



Here,

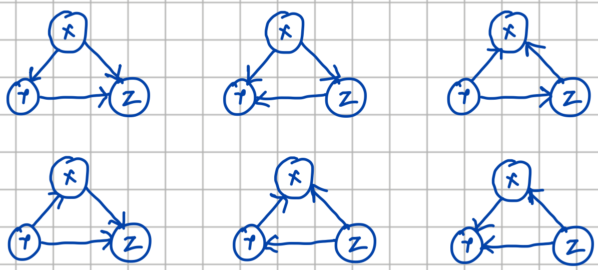
$X \not\perp\!\!\!\perp Y$ $\quad$ $X \not\perp\!\!\!\perp Y \mid Z$

$X \not\perp\!\!\!\perp Z$ $\quad$ $X \not\perp\!\!\!\perp Z \mid Y$

$Y \perp\!\!\!\perp Z$ $\quad$ $Y \not\perp\!\!\!\perp Z \mid X$

---

cluster 11 :



Here,

$X \not\perp\!\!\!\perp Y$ $\quad$ $X \not\perp\!\!\!\perp Y \mid Z$

$X \not\perp\!\!\!\perp Z$ $\quad$ $X \not\perp\!\!\!\perp Z \mid Y$

$Y \not\perp\!\!\!\perp Z$ $\quad$ $Y \not\perp\!\!\!\perp Z \mid X$

## Q3.

$$KL(p\|q) = -\int p(x) \ln\left(\frac{q(x)}{p(x)}\right) dx$$

$$= -\int p(x) \left(\ln q(x) - \ln p(x)\right) dx$$

$$= \int p(x) \left(\ln p(x) - \ln q(x)\right) dx$$

$$= \mathbb{E}\left[\ln p(x) - \ln q(x)\right] \qquad \left(\text{using } E[x]_{p(x)} = \int p(x) \cdot x \, dx\right)$$

on expanding terms of $p(x) = \mathcal{N}(x \mid \mu, \Sigma)$ & $q(x) = \mathcal{N}(x \mid m, L)$ :

$$= \mathbb{E}\left[\underbrace{-\frac{1}{2}\left(\cancel{D \ln 2\pi} + \ln|\Sigma| + (x-\mu)^T \Sigma^{-1} (x-\mu)\right)} + \underbrace{\frac{1}{2}\left(\cancel{D \ln 2\pi} + \ln|L| + (x-m)^T L^{-1} (x-m)\right)}\right]$$

$$= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} - \frac{1}{2} \mathbb{E}\left[(x-\mu)^T \Sigma^{-1} (x-\mu)\right] + \frac{1}{2} \mathbb{E}\left[(x-m)^T L^{-1} (x-m)\right]$$

Using (380) from matrix cookbook :

$$= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} - \frac{1}{2}\left[(\mu-\mu)^T \Sigma^{-1} (\mu-\mu) + Tr(\Sigma^{-1}\Sigma)\right] + \frac{1}{2}\left[(\mu-m)^T L^{-1} (\mu-m) + Tr(L^{-1}\Sigma)\right]$$

$$= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} - \frac{1}{2}\left[\vec{0} + \underbrace{Tr(I)}_{\substack{\text{sum of 1's} \\ \text{D times, where D} \\ \text{is dimen. of } x}}\right] + \frac{1}{2}\left[(\mu-m)^T L^{-1} (\mu-m) + Tr(L^{-1}\Sigma)\right]$$

$$= \frac{1}{2}\ln\frac{|L|}{|\Sigma|} - \frac{D}{2} + \frac{1}{2}\left[(\mu - m)^T L^{-1}(\mu - m) + Tr(L^{-1}\Sigma)\right]$$

Ans.

**Q4.**

Entropy of distribution 'p':

$$H(p) = -\int p(x) \ln p(x)\, dx$$

Here, $p(x) = \mathcal{N}(x|\mu, \Sigma)$. Substituting the definition of normal distribution in only the 'log' term we get:

$$= -\int -\frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + (x-\mu)^T \Sigma^{-1}(x-\mu) \cdot p(x)\, dx$$

$$= +\frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + \underline{\mathbb{E}\left[(x-\mu)^T\Sigma^{-1}(x-\mu)\right]} \qquad \left(\text{using } E[x]_{p(x)} = \int p(x)\cdot x\, dx\right)$$

Again, using (380) of matrix cook book:

$$= \frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + \left[(\mu-\mu)^T\Sigma^{-1}(\mu-\mu) + Tr(\Sigma^{-1}\Sigma)\right]$$

$$= \frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + \left[\vec{0} + Tr(\Sigma^{-1}\Sigma)\right]$$

$$= \frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + \underbrace{Tr(I)}_{= D \text{ (dimension of } x)}$$

$$= \frac{1}{2} D \ln 2\pi + \frac{1}{2}\ln|\Sigma| + \frac{D}{2}$$

Ans.