Changxin Miao
11853018
changxin.miao@student.uva.nl

Homework Assignment 2
Machine Learning 2, 17/18

2018-04-14

1 Problem 1

Mutual information

$$I(X;Y) = KL(p(X,Y) \| p(X)p(Y))) = H(X) - H(X \mid Y) = H(Y) - H(X \mid Y) \quad (1)$$

The mutual information is the difference of the information in X and the information left in X after we know Y. (Or the difference of the information in Y, and the information left in Y after we know X.)

$$I(X;Y \mid Z) = E_z(I(x;y) \mid Z) = \sum_{z \in Z} \sum_{y \in Y} \sum_{x \in X} p_z(z) p_{X,Y|Z} log \frac{p_{X,Y|Z}(X,Y \mid Z)}{p_{X|Z}(x \mid z) p_{Y|Z}(y \mid z)} \quad (2)$$

The conditional mutual information is the expected value of mutual information of two random variables given the value of the third one.

1.1 2

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

Tabelle 1: p(x,y) table

x	p(x)
0	0.6
1	0.4
y	p(y)
0	0.592
1	0.408

Tabelle 2: $p(x)$, $p(y)$ table

$$I(X,Y) = \sum_X \sum_Y p(X,Y) log(\frac{p(X,Y)}{p(X)p(Y)}) \quad (3)$$

$$I(X,Y) = 0.336 log \frac{0.336}{0.6 \times 0.592} + 0.264 log \frac{0.264}{0.6 \times 0.408} + 0.256 log \frac{0.256}{0.4 \times 0.592} + 0.144 log \frac{0.144}{0.4 \times 0.408} = 0.003197 \quad (4)$$

$I(X,Y) > 0$, it means that at least one pair of (x,y) has the property that the difference of information of x and the information left in x after knowing y is larger than 0. Therefore, the distribution of x and the distribution of y are not independent.

1.2 3

x	z	p(x,z)
0	0	0.24
0	1	0.36
1	0	0.24
1	1	0.16

Tabelle 3: p(x,z) table

y	z	p(y,z)
0	0	0.384
0	1	0.096
1	0	0.208
1	1	0.312

Tabelle 4: p(x,y) table

x	z	$p(x \mid z)$
0	0	0.5
0	1	0.096
1	0	0.208
1	1	0.312

Tabelle 5: $p(x \mid z)$ table

y	z	$p(y \mid z)$
0	0	0.8
0	1	0.4
1	0	0.2
1	1	0.6

Tabelle 6: $p(y \mid z)$ table

$$I(X;Y \mid Z) = \sum_X \sum_Y \sum_Z p(X,Y \mid Z) log(\frac{p(X,Y \mid Z)}{p(X \mid Z)p(Y \mid Z)}$$

$$I(X;Y \mid Z) = 0.4 log \frac{0.4}{0.5 \times 0.8} + 0.277 log \frac{0.277}{0.6923 \times 0.4} + 0.1 log \frac{0.1}{0.5 \times 0.2} + 0.4153 log \frac{0.4153}{0.6923 \times 0.6} + 0.4 log \frac{0.4}{0.5 \times 0.8}$$
$$+ 0.123 log \frac{0.123}{0.88 \times 0.4} + 0.1 log \frac{0.1}{0.5 \times 0.2} + 0.1846 log \frac{0.1846}{0.88 \times 0.6} = 0$$

x	y	z	$p(x, y \mid z)$
0	0	0	0.4
0	0	1	0.277
0	1	0	0.1
0	1	1	0.4153
1	0	0	0.4
1	0	1	0.123
1	1	0	0.1
1	1	1	0.1846

Tabelle 7: $p(x, y \mid z)$ table

This result implies that under the context of Z, the mutual information of X and Y is 0. In other words, the distribution of $(X \mid Z)$ and $(Y \mid Z)$ are independent.

1.3 4

As the distribution of $(X \mid Z)$ and $(Y \mid Z)$ are independent, we can infer that $p(X, Y \mid Z) = p(X \mid Z)p(Y \mid Z)$

$$\begin{aligned} p(X, Y, Z) &= p(X, Y \mid Z)p(Z) \\ &= p(X \mid Z)p(Y \mid Z)p(Z) \\ &= p(X \mid Z)p(Z)p(Y \mid Z) \\ &= p(Z \mid X)p(X)p(Y \mid Z) \end{aligned}$$

The directed graph will be displayed as below:

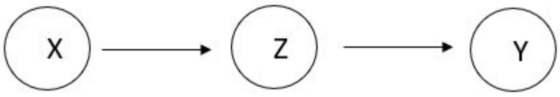

Abbildung 1: Directed graph for X,Y,Z

2 Problem 2

2.1 1

Fork case:

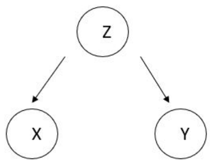

Abbildung 2: DAG for X,Y,Z

Independence relationship:
1. X⊥⊥Y|Z

$$p(X, Y \mid Z) = \frac{p(X, Y, Z)}{p(Z)} = \frac{p(X \mid Z)p(Y \mid Z)p(Z)}{p(Z)} = p(X \mid Z)p(Y \mid Z)$$

2. X$\not\perp\!\!\!\perp$Y

$$p(X, Y) = \sum_{Z} p(X, Y, Z) = \sum_{Z} p(X \mid Z)p(Y \mid Z)p(Z)$$

Compared with 1, the independence assumption does not hold.

2.2 2

Chain case:

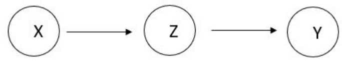

Abbildung 3: DAG for X,Y,Z

Independence relationship:
1.X $\perp\!\!\!\perp$ Y | Z

$$p(X, Y \mid Z) = \frac{p(X, Y, Z)}{p(Z)} = \frac{p(x)p(Z \mid X)p(Y \mid Z)}{p(Z)} = p(X \mid Z)p(Y \mid Z)$$

2.X $\not\perp\!\!\!\perp$ Y

$$p(X, Y) = \sum_{Z} p(Y \mid Z)p(Z \mid X)p(X) = p(Y \mid X)p(X) \neq p(Y)p(X)$$

2.3 3

Collide case:

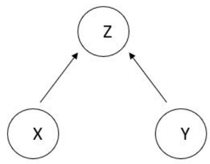

Abbildung 4: DAG for X,Y,Z

Independence relationship:
1.X $\perp\!\!\!\perp$ Y

$$p(X, Y) = \sum_{Z} p(X, Y, Z) = \sum_{Z} p(Z \mid X, Y)p(X)p(Y) = p(X)p(y)$$

2.X $\not\perp\!\!\!\perp$ Y | Z

$$p(X, Y \mid Z) = \frac{p(X, Y, Z)}{p(Z)} = \frac{p(Z \mid X, Y)p(X)p(Y)}{p(Z)} \neq p(X \mid Z)p(Y \mid Z)$$

3 Problem 3

3.1 1

$$
\begin{aligned}
KL(p\|q) &= \int p(x)ln\frac{q(x)}{p(x)}dx = -\int p(x)ln\frac{p(x)}{q(x)}dx \\
&= \int [ln\frac{1}{|\Sigma|^{\frac{1}{2}}} - ln\frac{1}{|L|^{\frac{1}{2}}} - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) + \frac{1}{2}(x-m)^T L^{-1}(x-m)]p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}\int Tr[(x-\mu)^T\Sigma^{-1}(x-\mu)]p(x)dx + \frac{1}{2}\int Tr[(x-m)^T L^{-1}(x-m)]p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}Tr[E(x-\mu)^T(x-\mu)]\Sigma^{-1}) + \frac{1}{2}\int [(x-m)^T L^{-1}(x-m)]p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{1}{2}Tr(\Sigma\Sigma^{-1}) + \frac{1}{2}\int [((x-\mu)+(\mu-m))^T L^{-1}((x-\mu)+(\mu-m))]p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{D}{2} + \frac{1}{2}\int [(x-\mu)^T L^{-1}(x-\mu) + 2(x-\mu)^T L^{-1}(\mu-m) + (\mu-m)^T L^{-1}(\mu-m)]p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{D}{2} + \frac{1}{2}Tr(\Sigma L^{-1}) + \int (x-\mu)^T L^{-1}(\mu-m)p(x)dx + \int (\mu-m)^T L^{-1}(\mu-m)p(x)dx \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{D}{2} + \frac{1}{2}Tr(\Sigma L^{-1}) + 0 + \frac{1}{2}(\mu-m)^T L^{-1}(\mu-m) \\
&= \frac{1}{2}ln\frac{|L|}{|\Sigma|} - \frac{D}{2} + \frac{1}{2}Tr(\Sigma L^{-1}) + \frac{1}{2}(\mu-m)^T L^{-1}(\mu-m)
\end{aligned}
$$

3.2 2

$$
\begin{aligned}
&\int p(x)dx = 1 \\
&\quad H(p) = -\int p(x)lnp(x)dx \\
&\qquad = -\int p(x)dx \times \int lnp(x)dx \\
&\qquad = -\int ln\frac{1}{(2\pi)^{\frac{D}{2}}} \mid \Sigma \mid^{-\frac{1}{2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))dx \\
&\qquad = \frac{D}{2}ln2\pi + \frac{1}{2}ln \mid \Sigma \mid + \frac{1}{2}\int Tr[(x-\mu)(x-\mu)^T\Sigma^{-1}] \\
&\qquad = \frac{D}{2}ln2\pi + \frac{1}{2}ln \mid \Sigma \mid + \frac{1}{2}Tr(E(x-\mu)(x-\mu)\Sigma^{-1}) \\
&\qquad = \frac{D}{2}ln2\pi + \frac{1}{2}ln \mid \Sigma \mid + \frac{D}{2}
\end{aligned}
$$