ASSESSMENT COVER SHEET

Assignment Details

Course: Machine Learning

Semester/Academic Year: Sem 2, 2015

Assignment title: Assignment 3

Assessment Criteria

Assessment Criteria are included in the Assignment Descriptions that are published on each course's web site.

Plagiarism and Collusion

Plagiarism: using another person's ideas, designs, words or works without appropriate acknowledgement.

Collusion: another person assisting in the production of an assessment submission without the express requirement, or consent or knowledge of the assessor.

Consequences of Plagiarism and Collusion

The penalties associated with plagiarism and collusion are designed to impose sanctions on offenders that reflect the seriousness of the University's commitment to academic integrity. Penalties may include: the requirement to revise and resubmit assessment work, receiving a result of zero for the assessment work, failing the course, expulsion and/or receiving a financial penalty.

DECLARATION

I declare that all material in this assessment is my own work except where there is clear acknowledgement and reference to the work of others. I have read the University Policy Statement on Plagiarism, Collusion and Related Forms of Cheating:

http://www.adelaide.edu.au/policies/?230

I give permission for my assessment work to be reproduced and submitted to academic staff for the purposes of assessment and to be copied, submitted and retained in a form suitable for electronic checking of plagiarism.

Zhao Xuanyu

SIGNATURE AND DATE 29/10, 2015.

Question1.

1. $P(A=a, B=b) = \sum_{C} P(a,b,C) = P(a,b,c) + P(a,b,\neg c)$

$= 0.02$

2. $P(B=b) = \sum_{A,C} P(A,b,C) = 0.01 + 0.01 + 0.04 + 0.04 = 0.1$

3. $P(A=a \mid B=b) = \dfrac{P(a,b)}{P(B=b)} = \dfrac{0.02}{0.1} = 0.2$

4. $P(A=\neg a \mid B=b) = 1 - P(A=a \mid B=b) = 0.8$

Question2.

1.

B	P(B)
1	1/3
2	1/3
3	1/3

B	P(apple|B)	P(orange|B)
1	1	0
2	0	1
3	$\frac{1}{2}$	$\frac{1}{2}$

2. $P(B=1, F=apple) = P(F=apple \mid B=1) \times P(B=1) = 1 \times \frac{1}{3} = \frac{1}{3}$.

3. $P(B=1 \mid F=apple) = \dfrac{P(F=apple \mid B=1) \cdot P(B=1)}{P(F=apple)} = \dfrac{\frac{1}{3}}{P(apple)}$.

$\because P(apple) = \sum_{B} P(B, apple) = \sum_{B} P(apple \mid B) \cdot P(B)$

$= 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} = \frac{1}{2}$.

$\therefore P(B=1 \mid F=apple) = \dfrac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$.

Question 3.

1. $P(G) = \sum_{D,I,S,L} P(D,I,G,S,L)$

$= \sum_{D,I,S,L} P(D)\,P(I)\,P(G|D,I)\,P(S|I)\,P(L|G)$

$= \sum_{D,I,S} P(D)P(I)P(G|D,I)P(S|I) \sum_{L} P(L|G)$

$= \sum_{D,I,S} P(D)P(I)P(G|D,I)\,P(S|I)$

$= \sum_{D,I} P(D)\,P(I)\,P(G|D,I) \sum_{S} P(S|I).$

$= \sum_{D,I} P(D)\,P(I)\,P(G|D,I)$

$= \sum_{D} P(D) \sum_{I} P(I)\,P(G|D,I)$

$= \sum_{D} P(D)\,m_1(G,D)$

$= m_2(G).$

2. Parameters are: $P(D)$, $P(I)$, $P(G|D,I)$, $P(S|I)$, $P(L|G)$.

Question 4.

$$\sum_{A} \hat{P}(A|B=0,C=0) = \sum_{A} \frac{N(A,B=0,C=0)+N_r}{N(B=0,C=0)+(\#A)\times N_r}$$

$$= \frac{\sum_{A} N(A,B=0,C=0) + \sum_{A} N_r}{N(B=0,C=0)+(\#A)\times N_r}$$

$\because \sum_{A} N_r = (\#A)\times N_r.$

$\because \sum_{A} N(A,B=0,C=0) = N(B=0,C=0)$

$$\therefore = \frac{N(B=0,C=0)+(\#A)\times N_r}{N(B=0,C=0)+(\#A)\times N_r} = 1. \quad \#$$

Question 5.

$\cancel{P(S) = \sum_{A,G} P}$ $P(S=1) = \dfrac{N(S=1)}{\sum_S N(S)} = \dfrac{4}{6} = \dfrac{2}{3}$,

$P(S=0) = 1 - P(S=1) = \frac{1}{3}$.

$\because \hat{p}(A|S) = \dfrac{N(A,S)+1}{N(S)+2\times 1}$, $\hat{p}(G|S) = \dfrac{N(G,S)+1}{N(S)+2\times 1}$.

$\therefore \hat{p}(A=0|S=0) = \dfrac{2+1}{2+2} = \dfrac{3}{4}$, $\hat{p}(A=1|S=0) = \dfrac{0+1}{2+2} = \dfrac{1}{4}$.

$\hat{p}(A=0|S=1) = \dfrac{2+1}{4+2} = \dfrac{1}{2}$, $\hat{p}(A=1|S=1) = \dfrac{2+1}{4+2} = \dfrac{1}{2}$.

$\therefore \hat{p}(G=0|S=0) = \dfrac{1+1}{2+2} = \dfrac{1}{2}$, $\hat{p}(G=1|S=0) = \dfrac{1+1}{2+2} = \dfrac{1}{2}$.

$\hat{p}(G=0|S=1) = \dfrac{3+1}{4+2} = \dfrac{2}{3}$, $\hat{p}(G=1|S=1) = \dfrac{1+1}{4+2} = \dfrac{1}{3}$.

Question 6.

1. $X \perp\!\!\!\perp Y | Z \Rightarrow P(X,Y|Z) = P(X|Z)P(Y|Z)$

$\because P(X,Y|Z) = P(Y,X|Z)$,

$\therefore P(Y,X|Z) = P(X|Z)P(Y|Z) = P(Y|Z) \cdot P(X|Z)$.

$\therefore Y \perp\!\!\!\perp X | Z$.

2. $\because X \perp\!\!\!\perp Y,W|Z$

$\therefore P(X,Y,W|Z) = P(X|Z) \cdot P(Y,W|Z)$.

$$\therefore \frac{P(X,Y,W,Z)}{P(Z)} = \frac{P(X,Z)}{P(Z)} \frac{P(Y,W,Z)}{P(Z)}$$

$$\therefore P(X,Y,W,Z) = \frac{P(X,Z)}{P(Z)} \cdot P(Y,W,Z). \qquad ①$$

Sum on W, we get

$$\sum_W P(X,Y,W,Z) = \frac{P(X,Z)}{P(Z)} \sum_W P(Y,W,Z)$$

$$\therefore P(X,Y,Z) = \frac{P(X,Z)}{P(Z)} \cdot P(Y,Z).$$

$$\Rightarrow \frac{P(X,Y,Z)}{P(Z)} = \frac{P(X,Z)}{P(Z)} \cdot \frac{P(Y,Z)}{P(Z)}.$$

$$P(X,Y|Z) = P(X|Z) \cdot P(Y|Z)$$

$$\therefore X \perp\!\!\!\perp Y | Z.$$

The same operation, sum on Y for formula ①.

$$\sum_Y P(X,Y,W,Z) = \frac{P(X,Z)}{P(Z)} \sum_Y P(Y,W,Z).$$

$$\Rightarrow \frac{P(X,W,Z)}{P(Z)} = \frac{P(X,Z)}{P(Z)} \cdot \frac{P(W,Z)}{P(Z)}$$

$$\Rightarrow P(X,W|Z) = P(X|Z) \cdot P(W|Z).$$

$$\Rightarrow X \perp\!\!\!\perp W | Z.$$

Question 7.

1. if X_i, X_j are independent.

for each (x_i, x_j), $P(x_i, x_j) = P(x_i)P(x_j)$

$\therefore I(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \cdot \log 1 = \sum_{x_i, x_j} P(x_i, x_j) \cdot 0 = 0$;

if $I(X_i, X_j) = 0$, and $P(x_i, x_j)$ is the joint probability.

$\Rightarrow P(x_i, x_j) > 0$.

$\because I(X_i, X_j) = \sum_{x_i, x_j} P(x_i, x_j) \cdot \log \frac{P(x_i, x_j)}{P(x_i)P(x_j)}$

$\therefore$ for any (x_i, x_j), we must have $\log \frac{P(x_i, x_j)}{P(x_i)P(x_j)} = 0$

$\Rightarrow \frac{P(x_i, x_j)}{P(x_i)P(x_j)} = 1 \quad \Rightarrow \quad X_i \perp\!\!\!\perp X_j$

2. if $P(X) = P'(X)$

$KL(P(X) \| P'(X)) = \sum_x P(x) \cdot \log \frac{P(x)}{P'(x)} = \sum_x P(x) \log 1 = 0$;

if $KL(P(X) \| P'(X)) = 0$, we get

$\sum_x P(x) \log \frac{P(x)}{P'(x)} = 0$,

$\because P(x) > 0$. $\therefore$ for any x, $\log \frac{P(x)}{P'(x)} = 0$

$\therefore \frac{P(x)}{P'(x)} = 1$

$\therefore$ for any x, $P(x) = P'(x)$.

$\Rightarrow P(X) = P'(X)$.

Question 8.

1. Since $k(\vec{u}, \vec{v})$ is a valid kernel.

$k(\vec{u}, \vec{v}) = \phi(\vec{u}) \cdot \phi(\vec{v})$. Thus the kernel matrix K_S should be

$(K_S)_{ij} = \phi(\vec{x_i}) \cdot \phi(\vec{x_j})$. Let $V = [\phi(x_1), \cdots \phi(x_n)]$, in the expression $\phi(x_i)$ is a column vector.

Then, $K_S = V^T V$.

Then, for any vector $\vec{z}$ in original space. we must ensure $(\vec{z}^T V^T)(V\vec{z}) \geq 0$. $\therefore$ $K_S = V^T V$ must be ≥ 0, must be PSD.

2. Dual form of binary SVM in soft margin case using kernel:

Maximize $L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \cdot K(\vec{x_i}, \vec{x_j})$.

subject to $0 \leq a_i \leq C$, $\sum_i a_i y_i = 0$.

3. ① Could map a non-linear problem to a high-dimension space, and in that space, the non-linear problem become a linear problem.

② Simple computation of the dot product in a potentially infinite dimensional feature space by means on the kernel function.

③ Simple construction of PSD kernels from other PSD kernels.

④ Given an algorithm formulated in terms of a PSD kernel k, we can formulate another algorithm by replacing k with another PSD kernel.