Solver: Fienny Angelina

Email Address: FIEN0001@e.ntu.edu.sg

1. (a)

(i) True. Deep learning is a type of methodologies of machine learning.

(ii) False. A and B are independent variables and each has direct influence on C.

(iii) False. Pre-pruning stops growing tree earlier.

(iv) False. It can go to local optima instead of global optimal

(v) False. KNN is a lazy learner

(vi) False. It is to maximize margin hyperplane

(vii) True. The precision is defined as $p = tp / (tp + fp)$, where the recall is defined as $r = tp/(tp + fn)$. If precision and recall are equal, we have $p = r$, and since they have the same denominator, we get $fp = fn$.

(viii) False. **Agglomerative**: This is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

(ix) True

(x) False. Principal components are constructed from the eigenvectors of the covariance

(b)

ID	X1	X2	Distance to (1,-2)
P1	2	2	4.12
P2	4	-3	3.16
P3	3	2	4.47
P4	-1	-4	2.83
P5	-5	2	7.21

Y = (YP1 + YP2+YP4)/3 = (10-2+4)/3 = 4

(c) For complex models, there is a greater chance that it was fitted accidentally by errors in data. Therefore given similar generalization errors, one should prefer simpler model.

2. (a) Let R(Am | X*) denotes the risk of taking action Y=m given X*
R(A1 | X*) = λ11 * P(Y=1| X*)+ λ12 * P(Y=2| X*) + λ13 * P(Y=3| X*) = 3.51
R(A2 | X*) = λ21 * P(Y=1| X*)+ λ22 * P(Y=2| X*) + λ23 * P(Y=3| X*) = 1.6
R(A3 | X*) = λ31 * P(Y=1| X*)+ λ32 * P(Y=2| X*) + λ33 * P(Y=3| X*) = 0.24
Since R(A3 | X*) is smallest, choose action Y=3.

(b)

(i)P(HC = Yes | C=High) = P(C=High | HC=Yes) P(HC=Yes) / P(C=High) …(1)

P(C=High | HC=Yes) = P(C=High, D=Healthy | HC=Yes) + P(C=High, D=Unhealthy | HC=Yes)

= P(C=High | D=Healthy) P(D=Healthy | HC=Yes) + P(C=High | D=Unhealthy) P(D=Unhealthy | HC=Yes)

= 0.2 * 0.8 + 0.7 * 0.2 = 0.3 ..(2)

P(HC = No | C=High) = P(C=High | HC= No) P(HC= No) / P(C=High) …(3)

P(C=High | HC= No) = P(C=High, D=Healthy | HC= No) + P(C=High, D=Unhealthy | HC= No)

= P(C=High | D=Healthy) P(D=Healthy | HC= No) + P(C=High | D=Unhealthy) P(D=Unhealthy | HC= No)

= 0.2 * 0.2 + 0.7 * 0.7 = 0.53 ..(4)

Substituting (2) to (1) and (4) to (3), and put it to (5)

P(HC = No | C=High) + P(HC=Yes | C=High) = 1

0.3 * 0.3 / P(C=High) + 0.53*0.7 / P(C=High) = 1. =====> P(C=High) = 0.461

P(HC=No | C=High) = 0.3 * 0.3 / 0.461 = 0.195

(ii)

P(D=Healthy, E=Yes) = P(D=Healthy, HC=Yes, E=Yes) + P(D=Healthy, HC=No, E=Yes)

= P(D=Healthy | HC=Yes, E=Yes) P(E=Yes, HC=Yes) + P(D=Healthy | HC=No, E=Yes) P(E=Yes, HC=No)

= 0.8 * (0.7 * 0.3) + 0.2 * (0.3 * 0.7) = 0.21……….(1)

P(D=Unhealthy, E=Yes) = P(D=Unhealthy, HC=Yes, E=Yes) + P(D=Unhealthy, HC=No, E=Yes)

= P(D=Unhealthy | HC=Yes, E=Yes) P(E=Yes, HC=Yes) + P(D=Unhealthy | HC=No, E=Yes) P(E=Yes, HC=No)

= 0.2 * (0.7 * 0.3) + 0.8 * (0.3 * 0.7) = 0.21 ………….(2)

P(D=Healthy | C=High, W=Normal, E=Yes) =

= P(D=Healthy, C=High, W=Normal, E=Yes) / P(C=High, W=Normal, E=Yes)

= P(C=High | D=Healthy, W=Normal, E=Yes) * P(D=Healthy, W=Normal, E=Yes) / P (C=High, W=Normal, E=Yes)

= P(C=High | D=Healthy, W=Normal, E=Yes) * P(W=Normal | D=Healthy, E=Yes) *P(D=Healthy, E=Yes) / P (C=High, W=Normal, E=Yes)

Substituting the values obtained from questions and (1):

P(D=Healthy | C=High, W=Normal, E=Yes) = 0.2 * 0.9 * 0.21 / P (C=High, W=Normal, E=Yes) …….(3)

Using similar method to simplify (3), we can denote that,

P(D=Unhealthy | C=High, W=Normal, E=Yes) =

= P(C=High | D=Unhealthy, W=Normal, E=Yes) * P(W=Normal | D=Unhealthy, E=Yes) *P(D=Unhealthy, E=Yes) / P (C=High, W=Normal, E=Yes)

P(D=Unhealthy | C=High, W=Normal, E=Yes) = 0.7 * 0.4 * 0.21 / P (C=High, W=Normal, E=Yes) ……..(4)

We also know that P(D=Unhealthy | C=High, W=Normal, E=Yes) + P(D=Healthy | C=High, W=Normal, E=Yes) = 1 …….(5)

Substitute (3) and (4) to (5)

0.2 * 0.9 * 0.21 / P (C=High, W=Normal, E=Yes) + 0.7 * 0.4 * 0.21 / P (C=High, W=Normal, E=Yes) = 1

P (C=High, W=Normal, E=Yes) = 0.0966 ………(6)

Substitute (3) to (6), we obtain:

P(D=Healthy | C=High, W=Normal, E=Yes) = 0.2 * 0.9 * 0.21 / P (C=High, W=Normal, E=Yes) = 0.2 * 0.9 * 0.21 /0.0966 = 0.391

3. (a) Initial Before Splitting

Yes	5
No	5

Entropy initial = - 0.5 log (0.5) - 0.5 * log (0.5) = 1

Splitting based on income

	Income < 5K	5K <= Income < 12K	Income > 12K
Yes	0	2	3
No	2	2	1

Entropy Income < 5K = 0
Entropy 5K<=Income<12K = 1
Entropy Income >12K = -0.75 log 0.75 - 0.25 log 0.25 = 0.8
Info gain = 1- (2/10) * 0 - (4/10) * 1 - (4/10) * 0.8
= 0.28

Splitting based on Occupation

	Teacher	Officer	Manager
Yes	3	0	2
No	1	3	1

Entropy Teacher = -0.75 log 0.75 - 0.25 log 0.25 = 0.8
Entropy Officer = 0
Entropy Manager = -(2/3) log(2/3) -(1/3) log(1/3) = 0.918

Info gain = 1 - (4/10) * 0.8 - 0 - (3/10) * 0.918
= 0.4046

(b)

(i)

$$w_{45}' = w_{45} - \lambda \frac{\partial E}{\partial w_{45}} = w_{45} - \lambda \boxed{\frac{\partial E}{\partial \hat{y}}} \cdot \boxed{\frac{\partial \hat{y}}{\partial z_5}} \cdot \boxed{\frac{\partial z_5}{\partial w_{45}}}$$

$(-1)(y-\hat{y})$

$z_5 = w_{35} h_3 + w_{45} h_4$

$\frac{\partial z_5}{\partial w_{45}} = h_4$

$\hat{y} = \text{sign}(z_5)$

$\frac{\partial \hat{y}}{\partial z_5} = 1$

$$= w_{45} + \lambda (y-\hat{y}) h_4$$

$$w_{24}' = w_{24} - \lambda \frac{\partial E}{\partial w_{24}} = w_{24} - \lambda \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_5} \cdot \boxed{\frac{\partial z_5}{\partial h_4}} \cdot \boxed{\frac{\partial h_4}{\partial z_4}} \cdot \boxed{\frac{\partial z_4}{\partial w_{24}}}$$

$z_5 = w_{35} h_3 + w_{45} h_4$

$\frac{\partial z_5}{\partial h_4} = w_{45}$

$h_4 = \text{sigmoid}(z_4)$

$\frac{\partial h_4}{\partial z_4} = (h_4)(1-h_4)$

$z_4 = w_{14} x_1 + w_{24} x_2$

$\frac{\partial z_4}{\partial w_{24}} = x_2$

$$w_{24}' = w_{24} + \lambda (y-\hat{y}) . w_{45} (h_4)(1-h_4) x_2$$

(ii) H3 = sigmoid(H1 * w13 + H2 * w23)

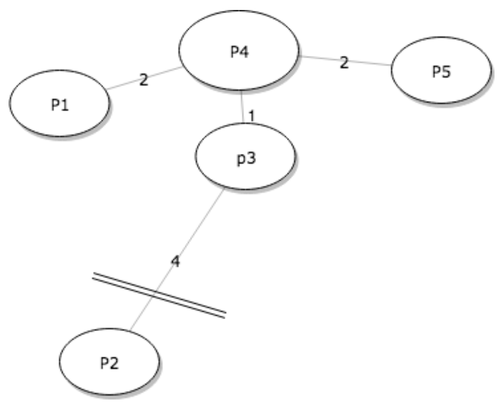
H4 = sigmoid(H1 * w14 + H2 * w24)

H5 = sign(H3 * w35 + H4 * w45)

Points	H1	H2	H3	H4	H5
P1	1	-1	0.881	0.5	1
P2	-1	0	0.269	0.3775	-1

Error rate = 0.5

4. (a)



(b)

(i) Try 3 entries with highest eigen value.
Pvar = (29.70 + 5.86 + 2.89) /(29.70 + 5.86 + 2.89 + 0.04) = 0.999 > 95%.
This is too large. Let’s try with 2 entries:
Pvar = (29.70 + 5.86) /(29.70 + 5.86 + 2.89 + 0.04) = 0.923
90% < 0.923 < 95%. Thus, k=2

(ii) The second and third line

(c)

(i) Try to find the number of points within 1.5 < 3 <= 4.5

There are 2 such points(P2, P6). p(3) = 2 / (6 * 3) = 0.1111

(ii) For Naive estimator, the interval (Δ) is fixed. However, for KNN estimator, the number of element (K) is fixed.

All the best for your exams!