Solver: Kenrick

Email Address: kenr0002@e.ntu.edu.sg

1.

$$P(disease = Yes) = 0.001$$
$$P(T_1 = +|disease = Yes) = 0.9$$
$$P(T_2 = +|disease = Yes) = 0.95$$
$$P(T_1 = +|disease = No) = 0.01$$
$$P(T_2 = +|disease = No) = 0.1$$

(a) $P(disease = Yes|T_1 = +, T_2 = +) = ?$

Using Bayes theorem,

$$P(disease = Yes|T_1 = +, T_2 = +) = \frac{P(disease = Yes, T_1 = +, T_2 = +)}{P(T_1 = +, T_2 = +)}$$
$$= \frac{P(T_1 = +, T_2 = +|disease = Yes) \times P(disease = Yes)}{P(T_1 = +, T_2 = +)}$$

Assuming T1 and T2 are independent,

$$P(T_1 = +, T_2 = +|disease = Yes)$$
$$= P(T_1 = +|disease = Yes) \times P(T_2 = +|disease = Yes) = 0.9 \times 0.95$$
$$= 0.855$$

Using sum rule,

$$P(T_1 = +, T_2 = +)$$
$$= P(T_1 = +, T_2 = +, disease = Yes) + P(T_1 = +, T_2 = +, disease = No)$$
$$= P(T_1 = +, T_2 = +|disease = Yes) \times P(disease = Yes)$$
$$+ P(T_1 = +, T_2 = +|disease = No) \times P(disease = No)$$
$$= P(T_1 = +|disease = Yes) \times P(T_2 = +|disease = Yes)$$
$$\times P(disease = Yes)$$
$$+ P(T_1 = +|disease = No) \times P(T_2 = +|disease = No)$$
$$\times P(disease = No) = 0.9 \times 0.95 \times 0.001 + 0.01 \times 0.1 \times (1 - 0.001)$$
$$= 1.854 \times 10^{-3}$$

Hence,

$$P(disease = Yes|T_1 = +, T_2 = +) = \frac{0.855 \times 0.001}{1.854 \times 10^{-3}} = 0.461$$

(b) Using 0-1 loss,

Risk of a_1 (predicting that there is disease) is:

$$R(a_1|T_1 = +, T_2 = +) = 1 - P(disease = yes|T_1 = +, T_2 = +) = 0.539$$

Risk of a_0 (predicting that there is no disease) is:

$$R(a_0|T_1 = +, T_2 = +) = 1 - P(disease = no|T_1 = +, T_2 = +)$$
$$= 1 - (1 - P(disease = yes|T_1 = +, T_2 = +)) = 0.461$$

Using the 0-1 loss, one should choose action that minimize the risk. Hence, it is predicted that the patient has no disease.

(c) Using cost/loss function defined,

Action	Actual	Cost
0	0	0
0	1	1
1	0	0.05
1	1	0

Risk of a_1 (predicting that there is disease) is:

$$R(a_1|T_1=+,T_2=+)=\sum_{k=0}^{1}\lambda_{1k}P(Y_k|T_1=+,T_2=+)$$
$$=\lambda_{10}P(disease=no|T_1=+,T_2=+)$$
$$+\lambda_{11}P(disease=yes|T_1=+,T_2=+)=0.05\times 0.539+0=0.0231$$

Risk of a_0 (predicting that there is no disease) is:

$$R(a_0|T_1=+,T_2=+)=\sum_{k=0}^{1}\lambda_{0k}P(Y_k|T_1=+,T_2=+)$$
$$=\lambda_{00}P(disease=no|T_1=+,T_2=+)$$
$$+\lambda_{01}P(disease=yes|T_1=+,T_2=+)=0+1\times 0.461=0.461$$

Using the defined loss function, one should choose action that minimize the risk. Hence, it is predicted that the patient has the disease.

---

2. (a)

X1 = 2, X2 = -0.5

n3 = sign(w13 * n1 + w23 * n3) = sign(0.5 * 2 + 1 * -0.5) = sign(0.5) = 1

n4 = sign(w14 * n1 + w24 * n2) = sign(0.5 * 2 + 1 * -0.5) = sign(0.5) = 1

n5 = sign(w35 * n3 + w45 * n4) = sign(0.5 * 1 + 1 * 1) = sign(2.5) = 1

(prediction) Y = +

Do similar process for all other classes, and you will get

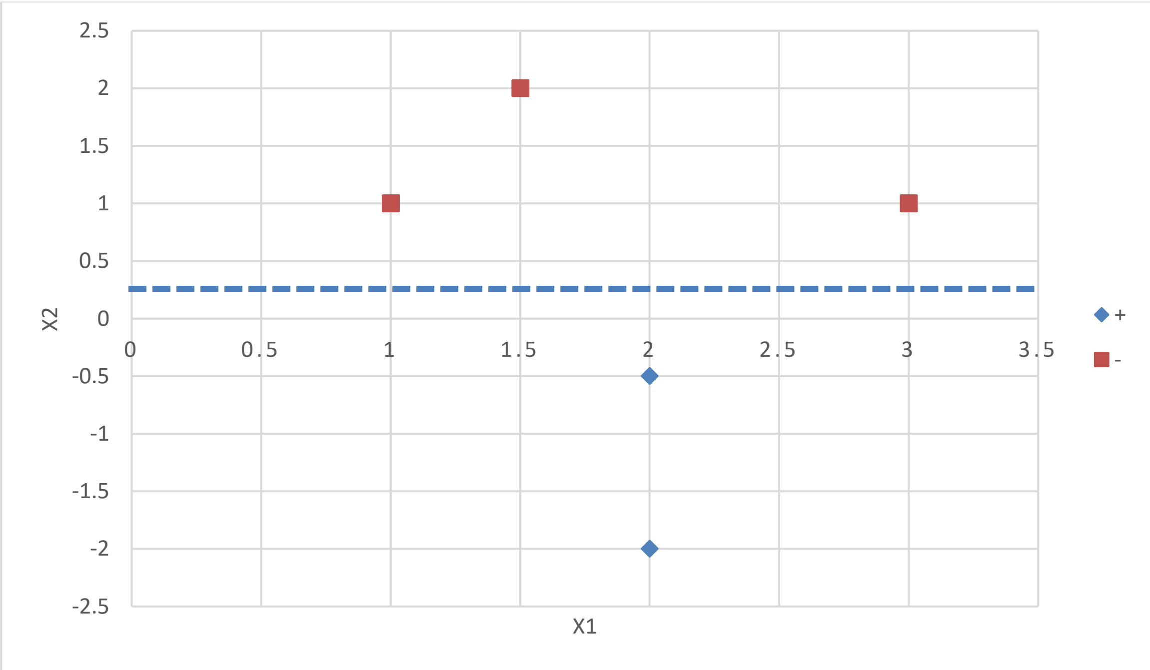
ID	Class label	Prediction
1	+	+
2	-	+
3	-	+
4	+	-
5	-	+

Error rate = 80%

(b)

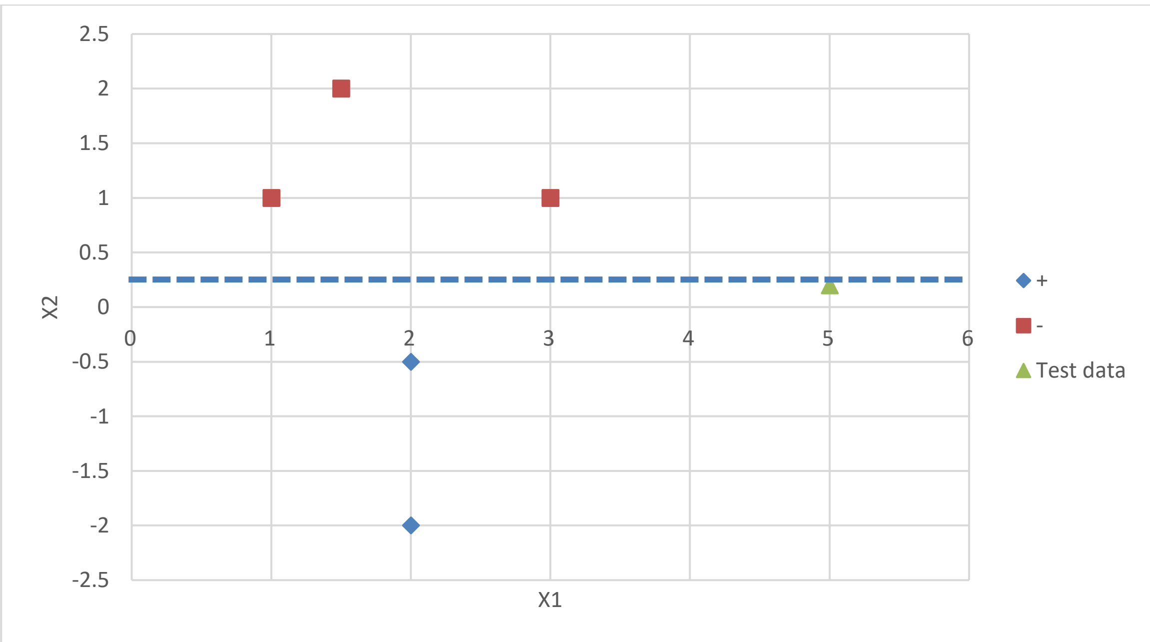
$$\begin{aligned}&\begin{pmatrix}2\\-0.5\end{pmatrix}\cdot\begin{pmatrix}2\\-0.5\end{pmatrix}+\begin{pmatrix}2\\-0.5\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}2\\-0.5\end{pmatrix}\cdot\begin{pmatrix}3\\1\end{pmatrix}+\begin{pmatrix}2\\-0.5\end{pmatrix}\cdot\begin{pmatrix}2\\-2\end{pmatrix}+\begin{pmatrix}2\\-0.5\end{pmatrix}\cdot\begin{pmatrix}1.5\\2\end{pmatrix}\\&+\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}3\\1\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}2\\-2\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1.5\\2\end{pmatrix}+\begin{pmatrix}3\\1\end{pmatrix}\cdot\begin{pmatrix}3\\1\end{pmatrix}+\begin{pmatrix}3\\1\end{pmatrix}\cdot\begin{pmatrix}2\\-2\end{pmatrix}\\&+\begin{pmatrix}3\\1\end{pmatrix}\cdot\begin{pmatrix}1.5\\2\end{pmatrix}+\begin{pmatrix}2\\-2\end{pmatrix}\cdot\begin{pmatrix}2\\-2\end{pmatrix}+\begin{pmatrix}2\\-2\end{pmatrix}\cdot\begin{pmatrix}1.5\\2\end{pmatrix}+\begin{pmatrix}1.5\\2\end{pmatrix}\cdot\begin{pmatrix}1.5\\2\end{pmatrix}=49.5\end{aligned}$$

(c)



At the chart, draw a boundary line that maximally separates (+) and (-) data points. This will be the decision boundary of the linear SVM. From this chart, the decision boundary is line $x_2 = 0.25$

(d)



After plotting it to the chart, it is seen that the test data point is on the decision area of (+), hence it is predicted that this test data point is (+)

3. (a)

Eigenvalue	Percentage
9.78	$\frac{9.78}{9.78+2.11+0.11} = 81.5\%$
2.11	$\frac{2.11}{9.78+2.11+0.11} = 17..58\%$
0.11	$\frac{0.11}{9.78+2.11+0.11} = 0.916\%$

(b) Note that the data points given in Table Q3 need to do a mean-correction but they have been mean-corrected.

$\begin{pmatrix}2\\0\\2\end{pmatrix}$	$\begin{pmatrix}0.78 & -0.54 & 0.32\\0.04 & 0.56 & 0.83\end{pmatrix}\begin{pmatrix}2\\0\\2\end{pmatrix} = \begin{pmatrix}2.2\\1.74\end{pmatrix}$
$\begin{pmatrix}-1\\0\\-1\end{pmatrix}$	$\begin{pmatrix}0.78 & -0.54 & 0.32\\0.04 & 0.56 & 0.83\end{pmatrix}\begin{pmatrix}-1\\0\\-1\end{pmatrix} = \begin{pmatrix}-1.1\\0.87\end{pmatrix}$
$\begin{pmatrix}3\\3\\0\end{pmatrix}$	$\begin{pmatrix}0.78 & -0.54 & 0.32\\0.04 & 0.56 & 0.83\end{pmatrix}\begin{pmatrix}3\\-3\\0\end{pmatrix} = \begin{pmatrix}3.96\\-1.56\end{pmatrix}$
$\begin{pmatrix}-3\\2\\-2\end{pmatrix}$	$\begin{pmatrix}0.78 & -0.54 & 0.32\\0.04 & 0.56 & 0.83\end{pmatrix}\begin{pmatrix}-3\\2\\-2\end{pmatrix} = \begin{pmatrix}-4.06\\-0.66\end{pmatrix}$
$\begin{pmatrix}-1\\1\\1\end{pmatrix}$	$\begin{pmatrix}0.78 & -0.54 & 0.32\\0.04 & 0.56 & 0.83\end{pmatrix}\begin{pmatrix}-1\\1\\1\end{pmatrix} = \begin{pmatrix}-1\\1.35\end{pmatrix}$

(c) Single-linkage (MIN) criterion.

Distance	x1	x2	x3	x4	x5
x1	0.00	3.41	3.74	6.70	3.22
x2	3.41	0.00	5.61	3.33	0.49
x3	3.74	5.61	0.00	8.07	5.75
x4	6.70	3.33	8.07	0.00	3.66
x5	3.22	0.49	5.75	3.66	0.00

Merge x2 and x5

Distance	x1	x2, x5	x3	x4
x1	0.00	3.22	3.74	6.70
x2, x5	3.22	0.00	5.61	3.33
x3	3.74	5.61	0.00	8.07
x4	6.70	3.33	8.07	0.00

Merge {x2, x5} and x1

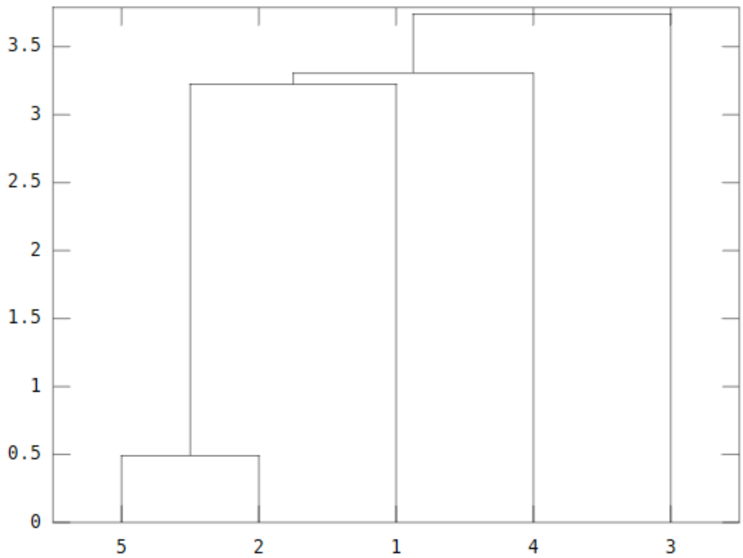
Distance	x2, x5, x1	x4	x3
x2, x5, x1	0.00	3.74	3.33
x4	3.74	0.00	8.07
x3	3.33	8.07	0.00

Merge {x2, x5, x1} and x4

Distance	x2, x5, x1, x4	x3
x2, x5, x1, x4	0.00	3.74
x3	3.74	0.00

Merge {x2, x5, x1, x4} and x3

Dendrogram:



(d) Feature embedding uses the covariance matrix $\mathbf{XX}^T$ (which are N-dimensions) instead of PCA's $\mathbf{X}^T\mathbf{X}$ (d-dimensions) to find the k eigenvectors with largest eigenvalues to project the N data points to k dimensions.

4. (a) I'm not too sure on this part, especially (vi) and (ix)
   - i) Split D into D1 and D2
   - ii) C1 train with D1
   - iii) C1 test with D2
   - iv) Error samples from (ii) + equal count of correct samples are taken as D3
   - v) C2 train with D3
   - vi) C2 test with D3
   - vii) Error samples from (vi) + equal count of correct samples are taken as D4
   - viii) C3 train with D4
   - ix) C3 test with D4

   (b)

Points in class 1	Distance to point	Points in class 2	Distance to point
(-1, 4)	4.123	(-1, 1)	1.414
(-1, 0)	1	(2, 1)	2.236
(1, -1)	1.414	(3, -2)	3.606

When k=3, points considered are (-1,0), (1, -1), (-1, 1); there are two in class 2 and one in class 1.

$$P(class = 1|x) = \frac{k_1}{k} = \frac{1}{3}$$
$$P(class = 2|x) = \frac{k_2}{k} = \frac{2}{3}$$

When k=5, points considered are (-1,0), (1, -1), (-1, 1), (2,1), (3, -2); there are three in class 2 and two in class 1.

$$P(class = 1|x) = \frac{k_1}{k} = \frac{2}{5}$$
$$P(class = 2|x) = \frac{k_2}{k} = \frac{3}{5}$$

(c) (i)

$$\begin{aligned}P(A = N) &= \sum_i \sum_j P(A = N, B = i, E = j)\\
&= P(A = N, B = Y, E = Y) + P(A = N, B = Y, E = N)\\
&+ P(A = N, B = N, E = Y) + P(A = N, B = N, E = N)\\
&= P(A = N|B = Y, E = Y) \times P(B = Y) \times P(E = Y)\\
&+ P(A = N|B = Y, E = N) \times P(B = Y) \times P(E = N)\\
&+ P(A = N|B = N, E = Y) \times P(B = N) \times P(E = Y)\\
&+ P(A = N|B = N, E = N) \times P(B = N) \times P(E = N)\\
&= 0.05 \times 0.01 \times 0.02 + 0.10 \times 0.01 \times 0.98 + 0.75 \times 0.99 \times 0.02\\
&+ 0.99 \times 0.99 \times 0.98 = 0.976\end{aligned}$$

(ii)

$$\begin{aligned}P(S = Y) &= \sum_i P(S = Y, A = i) = \mathrm{P}(\mathrm{S} = \mathrm{Y}, \mathrm{A} = \mathrm{Y}) + \mathrm{P}(\mathrm{S} = \mathrm{Y}, \mathrm{A} = \mathrm{N})\\
&= \mathrm{P}(\mathrm{S} = \mathrm{Y}|\mathrm{A} = \mathrm{Y}) \times \mathrm{P}(\mathrm{A} = \mathrm{Y}) + \mathrm{P}(\mathrm{S} = \mathrm{Y}|\mathrm{A} = \mathrm{N}) \times \mathrm{P}(\mathrm{A} = \mathrm{N})\\
&= 0.70 \times (1 - 0.976) + 0.05 \times 0.976 = 0.0656\end{aligned}$$

(iii)

$$P(B|S) = \frac{P(S|B)P(B)}{P(S)}$$

$$P(S|B) = \sum_{A,E} P(S|A)P(A|B,E)P(E)P(B)$$
$$= 0.631$$
$$P(B|S) = 0.097$$

My personal course "cheatsheets" can be obtained at: blog.kenrick95.org/resources

For reporting of errors and errata, please visit pypdiscuss.appspot.com
Thank you and all the best for your exams! ☺