Indian Institute of Information Technology Sri City

EndSem; **Machine Learning***; 2019; For 35 Marks*

Calculators are allowed; closed book & closed notes exam;180 min. duration.

Assume any missing information appropriately by clearly stating them.

1	**For a two class two dimensional classification problem, assume that the class conditional densities are Gaussian. Let the mean and covariance matrix for the class ω_1 be $\mu_1 = \begin{pmatrix}0\\0\end{pmatrix}$, $\Sigma_1 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ and that for the class ω_2 be $\mu_2 = \begin{pmatrix}4\\4\end{pmatrix}$, $\Sigma_2 = \begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}$. Find the decision surface when we use the Bayes classifier. Pictorially can you show this as a 2D plot? Assume that the prior probabilities are, $P(\omega_1) = \frac{3}{4}$ and $P(\omega_2) = \frac{1}{4}$. [6 Marks]**

Multivariate Normal distribution $p(X) \sim N(\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\lvert\Sigma\rvert^{1/2}} \exp\left[-\frac{1}{2}(X-\mu)^t\Sigma^{-1}(X-\mu)\right]$

Log of the posterior probability (after cancelling irrelevant things) gives, the discriminant function for class ω_1

$$g_1(X) = -\frac{1}{2}(X-\mu_1)^t\Sigma_1^{-1}(X-\mu_1) - \frac{1}{2}\ln\lvert\Sigma_1\rvert + \ln P(\omega_1)$$
$$= -\frac{1}{2}\left(x_1^2 + x_2^2\right) - \frac{1}{2} + \ln 3 - \ln 4$$

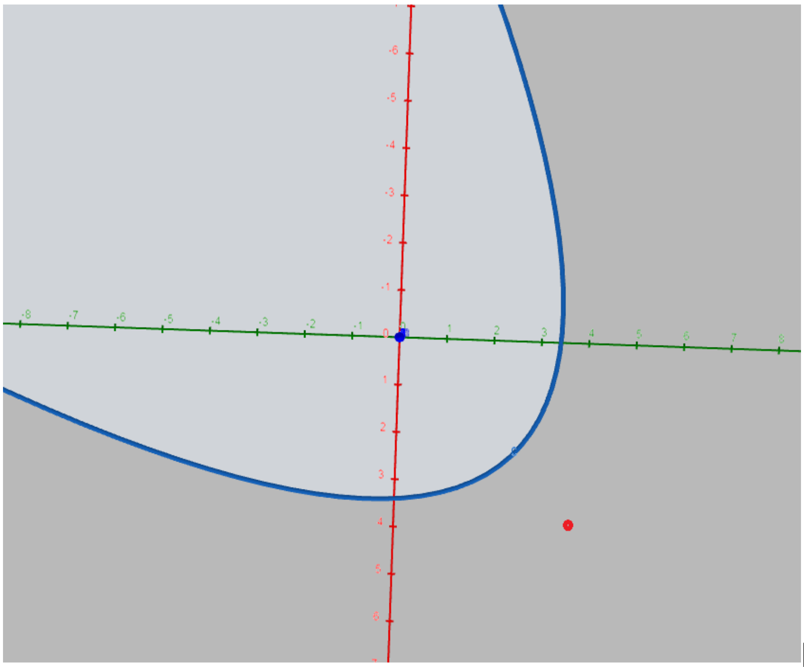
To find $g_2(X)$, we need Σ_2^{-1} *and* $\lvert\Sigma_2\rvert$:

$\Sigma_2^{-1} = \frac{1}{3}\begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}$, and $\lvert\Sigma_2\rvert = 3$. We get

$$g_2(X) = -\frac{1}{3}\left(x_1^2 + x_2^2 - x_1x_2 - 4x_1 - 4x_2 + 16\right) - \frac{1}{2}\ln 3 - \ln 4$$

Decision surface is $g(X) = g_1(X) - g_2(X)$

$= -x_1^2 - x_2^2 + 2x_1x_2 - 8x_1 - 8x_2 + 29 + 9\ln 3 = 0$[4 Marks]



[2 Marks]

2 **Consider a two dimensional (input-space with co-ordinates $(x_1, x_2)^t$) two class problem. Training set for $+1$ class is $\{X_1 = (1,1)^t, X_2 = (5,5)^t, X_3 = (6,6)^t\}$; that for -1 class is $\{X_4 = (8,8)^t, X_5 = (12,12)^t\}$. Find the hard linear SVM classifier. Find the margin of the classifier. Show pictorially your solution. [Trial and error solution will get zero, if not negative marks. You need to follow the procedures described in the class.][4+2=6 Marks]**

From geometry we get, $\alpha_1 = \alpha_2 = \alpha_5 = 0$; and also we get $\alpha_3 = \alpha_4$
Let us call $\alpha_3 = \alpha_4 = \alpha$
From KKT conditions, we get

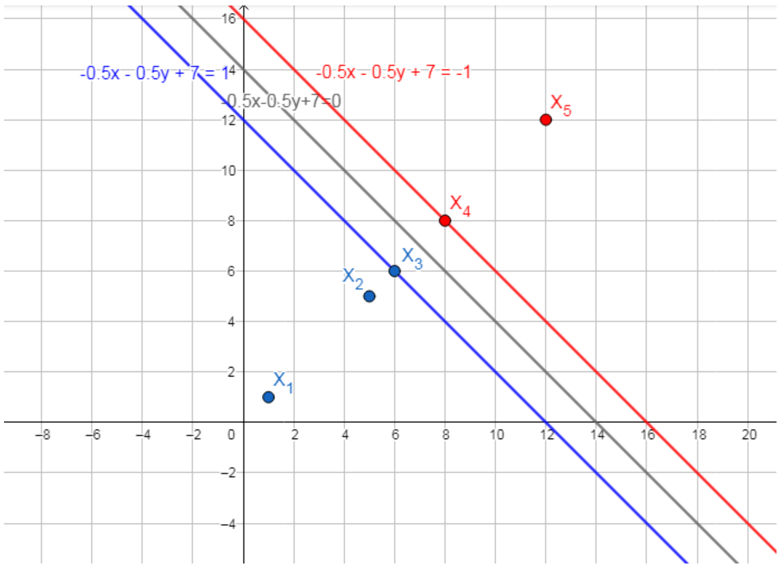
$$W = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \sum_{i=1}^{5} \alpha_i y_i X_i = \begin{pmatrix} -2\alpha \\ -2\alpha \end{pmatrix}$$

$$1 - (1)\left[\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 6 \end{pmatrix} + b\right] = 0 - - - \rightarrow \ (1)$$

$$1 - (-1)\left[\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \cdot \begin{pmatrix} 8 \\ 8 \end{pmatrix} + b\right] = 0 - - - \rightarrow (2)$$

Solving the above by replacing $w_1 = w_2 = -2\alpha$, we get $W = \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}$, and $b = 7$

[4 Marks]



Margin $= \frac{2}{\|W\|} = 2\sqrt{2}$[2 Marks]

3	One researcher tried to solve the XOR problem by using the logistic regression classifier, but he failed. His idea is that since XOR requires a non-linear solution, the sigmoid activation function present in the logistic regression, which is non-linear, can handle this. One of his friends, who apparently a student of IIITS, said that this can be done, but one has to work in a new space. The friend's idea is to map the 2D input space in to a new space, by using the mapping: $(x_1, x_2)^t \to (1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^t$ and then to learn the logistic regression classifier. Let the data that belongs to the class $+1$ be $\left\{X_1 = \binom{0}{0},\ X_2 = \binom{1}{1}\right\}$,and that belongs to the class 0be $\left\{X_3 = \binom{1}{0},\ X_4 = \binom{0}{1}\right\}$. Let the solution in the new space (after applying the mapping) be$y = 1/(1 + e^{-g(X)})$ where $g(X) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_1^2 + \theta_4x_2^2 + \theta_5x_1x_2$. (a) By using the log-likelihood criterion (as discussed in the class) solve this problem with the starting initial guess being$(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (1, 1, 1, 1, 1, 1)$ for 2 iterations using the ***batch*** method of learning.You need to give appropriate intermediate steps. Learning rate you can use is 0.1. (b) As per your classifier what would be the outputs, viz., y_1, y_2, y_3, y_4 for the patterns X_1, X_2, X_3, X_4. (c) Can you give a two line reason why the original idea, of applying logistic regression in the input spacefailed?[5+2+1 = 8Marks]
	Let the mapping be $\phi(X)$. The points after the mapping would be, $\phi(X_1) = (1, 0, 0, 0, 0, 0)^t$ $\phi(X_2) = (1, 1, 1, 1, 1, 1)^t$ $\phi(X_3) = (1, 1, 0, 1, 0, 0)^t$ $\phi(X_4) = (1, 0, 1, 0, 1, 0)^t$ 1st iteration: $g(X_1) = 1, so\ we\ get\ y_1 = 0.7311$ we have $t_1 = +1$ $g(X_2) = 6, so\ we\ get\ y_2 = 0.9975;\ \ t_2 = +1$ $g(X_3) = 3, so\ we\ get\ y_3 = 0.9526;\ \ t_3 = 0$ $g(X_4) = 3, so\ we\ get\ y_4 = 0.9526;\ \ t_4 = 0$ $$\theta_{new} = \theta + \eta \sum_{i=1}^{n} (t_i - y_i)\, \phi(X_i)$$ $$\theta_{new} = \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} + 0.1\left(0.7811\begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix} + 0.0025\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} - 0.9526\begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix} - 0.9526\begin{bmatrix}1\\0\\1\\0\\1\\0\end{bmatrix}\right)$$ $$\theta_{new} = \begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix} + 0.1\begin{bmatrix}-1.6337\\-0.9501\\-0.9501\\-0.9501\\-0.9501\\0.0025\end{bmatrix} = \begin{bmatrix}0.8366\\0.9050\\0.9050\\0.9050\\0.9050\\1.0002\end{bmatrix}$$

[5 Marks]

2nd iteration:

$g(X_1) = 0.8366$, $so\ we\ get\ y_1 = 0.6978$ **we have** $t_1 = +1$
$g(X_2) = 5.4568, so\ we\ get\ y_2 = 0..9958;\ \ t_2 = +1$
$g(X_3) = 2.6466, so\ we\ get\ y_3 = 0.9338;\ \ t_3 = 0$
$g(X_4) = 2.6466, so\ we\ get\ y_4 = 0.9338;\ \ t_4 = 0$

$$\theta_{new} = \begin{bmatrix} 0.8366 \\ 0.9050 \\ 0.9050 \\ 0.9050 \\ 0.9050 \\ 1.0002 \end{bmatrix} + 0.1 \begin{bmatrix} -1.5611 \\ -0.9296 \\ -0.9296 \\ -0.9296 \\ -0.9296 \\ 0.0042 \end{bmatrix} = \begin{bmatrix} 0.6805 \\ 0.8120 \\ 0.8120 \\ 0.8120 \\ 0.8120 \\ 1.0007 \end{bmatrix}$$

After the 2nd iteration:

$g(X_1) = 0.6805;\ g(X_2) = 4.9293;\ g(X_3) = 2.3046;\ g(X_4) = 2.3046;$

$$y_1 = 0.6639;\ y_2 = 0.9928;\ y_3 = 0.9093;\ y_4 = 0.9093;$$

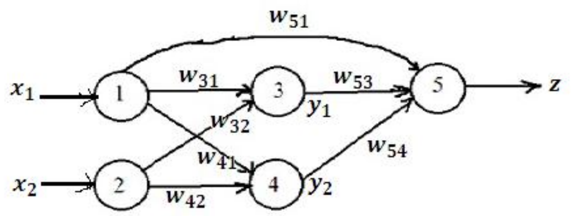
[2 Marks]

Original ideal fails because logistic regression finds a discriminant which is linear.

[1 Mark]

4

For the shown neural network where nodes 1 and 2 are the input neurons and node 5 is the output neuron. It is a conventional feed-forward network except that there is a connection from node 1 to node 5. Let the given input is $(x_1, x_2)^t = (1, 1)^t$ and the target output $t = 0$. Let all weights are currently equal to 1. Consider the sigmoid activation function $f(v) = \frac{1}{1+e^{-v}}$, and the learning rate 0.1. Consider the error, E = $\frac{1}{2}(t-z)^2$. Note, there is no bias unit present in the network.
(a) Find the weight updation rule for each of the weights.
(b) Find the updated weights after one iteration.
(c) Pictorially draw the network, both before and after weight updation by showing weights and output.



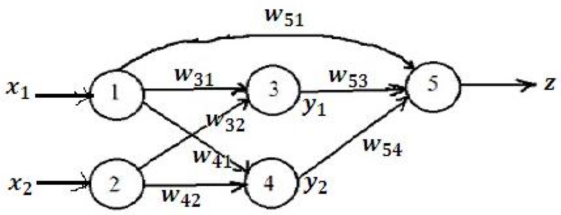
{ Use the notation given, which is, weight from node i to node j is w_{ji} . Inputs are $x_1,\ x_2$, output of nodes 3, 4 are y_1, y_2.
Output is of the network is . Target is t }

[4+2+2 = 8 Marks]

$Y_1 = 1/(1+e^{-2})$ = 0.88079, here net_3 or v3 = 2
$Y_2 = 1/(1+e^{-2})$ = 0.88079, here net_4 or v_4 = 2
$Z = 1/(1+e^{-2.76158})$ = 0.94056, here net_5 or v_5 = 2
Given error function J = $\frac{1}{2}(t-z)^2$
We know that Z = f(net_5),
Where $net_5 = y_1 w_{53} + x_1 w_{51} + y_2 w_{54}$ = 2.76158
$Y_1 = f(net_3)$ =0.88079
$net_3 = x_1 w_{31} + x_2 w_{32} = 2$
$Y_2 = f(net_4)$ = 0.88079
$Net_4 = x_1 w_{41} + x_2 w_{42} = 2$
The parameters (weights) of the network can be updated using gradient descent as follows,
$w_{new} = w_{old} + \eta(-\nabla J w_{old})$
Using chain rule,
$W_{51new} = w_{51old} + \eta\, [(t-z)f(net_5)(1-f(net_5)x_1]$
$W_{53new} = w_{53old} + \eta\, [(t-z)f(net_5)(1-f(net_5)y_1]$
$W_{54new} = w_{54old} + \eta\, [(t-z)f(net_5)(1-f(net_5)y_2]$
$W_{31new} = w_{31old} + \eta\, [(t-z)f(net_5)(1-f(net_5)w_{53}f(net_3)(1-f(net_3)x_1]$
$W_{32new} = w_{32old} + \eta\, [(t-z)f(net_5)(1-f(net_5)w_{53}f(net_3)(1-f(net_3)x_2]$
$W_{41new} = w_{41old} + \eta\, [(t-z)f(net_5)(1-f(net_5)w_{54}f(net_4)(1-f(net_4)x_1]$
$W_{42new} = w_{42old} + \eta\, [(t-z)f(net_5)(1-f(net_5)w_{54}f(net_4)(1-f(net_4)x_2]$
[4 Marks]

By Substituting the required values in the above equations, we get
W_{51} = 0.994741
W_{53} = 0.995368
W_{54} = 0.995368
W_{31} = 0.999447
W_{32} = 0.999447

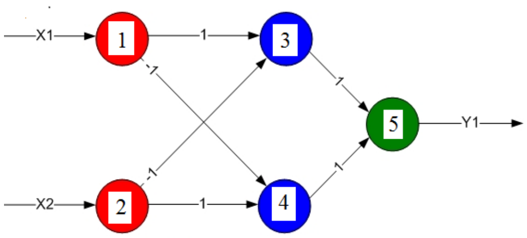
W_{41} = 0.999447
W_{42} = 0.999447
[2 Marks]



Before updating the weights, the values are
W_{51} = 1
W_{53} = 1
W_{54} = 1
W_{31} = 1
W_{32} = 1
W_{41} = 1
W_{42} = 1
Y_1 = 0.88079
Y_2 = 0.88079
Z = 0.94056

After updating the weights, the corresponding values are
W_{51} = 0.994741
W_{53} = 0.995368
W_{54} = 0.995368
W_{31} = 0.999447
W_{32} = 0.999447
W_{41} = 0.999447
W_{42} = 0.999447
Y_1 = 0.88068
Y_2 = 0.88068
Z = 0.93979
[2 Marks]

5

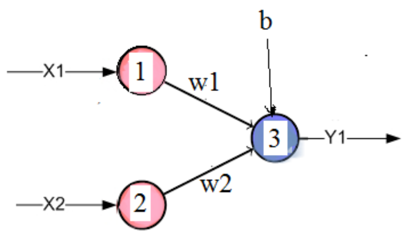


(Weights connecting 1 and 4, as well as 2 and 3 are -1; but all other weights are 1.)

---- --

One person while trying to solve the XOR problem found the above shown multilayer perceptron. But, that person wrongly gave

Since activation function used is linear, the multilayer perceptron is equal in power to a single layer perceptron. like the following one.



Assuming that the activation function for the node 3 (in this second figure) is $f(t) = t$, can you findout the weights w1,

the activation function as $f(t) = 2t + 1$, for nodes 3, 4 and 5.	w2 and the bias b so that the two neural networks are equivalent in their outputs. [3 Marks]

From the first figure:

$net_3 = x1 - x2$, **So output of node 3 will be** $o_3 = 2(x1 - x2) + 1$

$net_4 = -x1 + x2$, **So output of node 4 will be** $o_4 = 2(-x1 + x2) + 1$

$$net_5 = 2x1 - 2x2 + 1 - 2x1 + 2x2 + 1 = 2,$$

So output of node 5 will be $y1 = 2 \times 2 + 1 = 5$.**[2 Marks]**

Observe that y1 is independent of x1 and x2;

So in the second figure, we get w1 = 0; w2 = 0; and b = 5; [1 Mark]

---

6 **The given data in 3D Euclidean space is** $D = \{X_1 = (2,3,4)^t, X_2 = (3,4,5)^t, X_3 = (4,5,6)^t, X_4 = (5,6,7)^t, X_5 = (6,7,8)^t\}$. **As per the Principal Component Analysis, find the best 1D representation of the data. That is, your answer should be** $\{a_1, a_2, a_3, a_4, a_5\}$ **where each one is a scalar.**

The representation you got is loss-less or lossy? Justify your answer. [3+1 = 4Marks]

Mean vector = $\mu = (4,5,6)^t$

Mean subtracted data will be $\{(-2,-2,-2)^t, (-1,-1,-1)^t, (0,0,0)^t, (1,1,1)^t, (2,2,2)^t\}$

Scatter Matrix = $S = (-2,-2,-2)(-2,-2,-2)^t + \cdots + (2,2,2)(2,2,2)^t = \begin{bmatrix} 10 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 10 \end{bmatrix}$

Eigen values would be 30, 0, 0

Eigen vector corresponding to the largest eigen value would be $(1,1,1)^t$

But we need unit vector, so $e_1 = \frac{1}{\sqrt{3}}(1,1,1)$

Projection Matrix P = $\frac{1}{\sqrt{3}}(1,1,1)$

So, $a_1 = P(-2,-2,-2)^t = -6/\sqrt{3}$

$$a_2 = P(-1,-1,-1)^t = -3/\sqrt{3}$$
$$a_3 = 0$$
$$a_4 = 3/\sqrt{3}$$
$$a_5 = 6/\sqrt{3}$$

Same answer can be obtained from covariance matrix also,

$$\Sigma = \frac{1}{5}S = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

Eigen values of the covariance matrix would be 6, 0 and 0.

Eigen vector would be same as that of the scatter matrix.

So P also will be same, hence the final answer will be same. [3 Marks]

The representation is loss-less. Both the lower eigen values are zero. This means we can recover the data as it is from the 1D representation. This also means that the given data indeed is a 1D data (all 5 points are on the same straight line). [1 Mark]

--the end--