Solution and Marking Scheme of Machine Learning

Q.1 the problem is overfitting. (2.75 mark for identification)

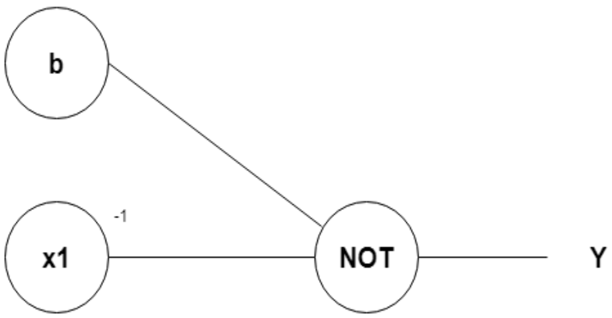
Solutions to overfitting (any 3) (short description is required for all three):(2 marks for each correct explanation)

1. Train with more data
2. Regularization
3. Feature selection
4. Changing complexity (number and value of weights) of network

(One for correct identification and one for correct explanation)

1. Classification, as output is only yes or no (2 marks)
2. Regression , as output is continuous variable (2 marks)
3. Regression , as relationship of CEO salary based on variables is to be predicted(2 marks)
4. Regression , as output is continuous (2 marks)
5. Classification, as image can be from one of the specified classes. (2 marks)

Q.2 (total 6 marks (3 for designing + explanation and 3 for correct calculation))



pick *w = -1* and *b = 0.5* *(or any other possible bias and weight giving correct answer)*

when X1=1 :

1 * (-1) + 0.5= -0.5 < 0 so output =0 (after application of heavyside step function as the activation function)

When X1= 0 :

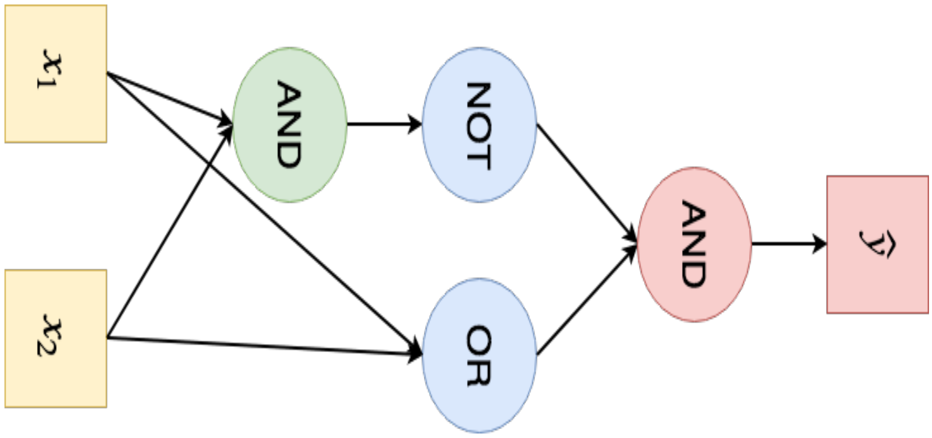
0 * -1 + 0.5 = 0.5 >0 so output = 1 (after application of heavyside step function as the activation function)

XOR function cannot be modelled by using single layer perceptron because it is not linearly separable. (2 marks)

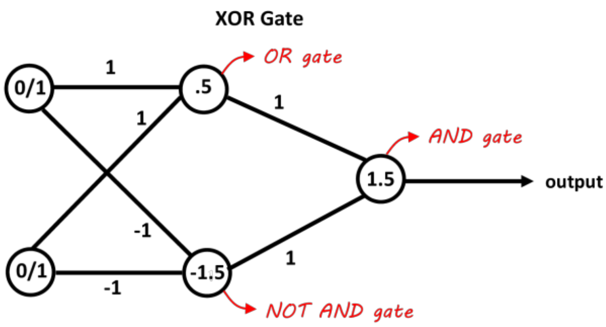
(or any other correct function is valid)

Solution is to use multilayer neural networks as shown below. (Explanation of the solution + calculation demonstrating correct results is required) (4.75 marks)

$$XOR(x1, x2) = AND(NOT(AND(x1, x2)), OR(x1, x2))$$



Sample weights could be :



6 marks for the below :

X= 08* 0.2 +0.6* 0.1 +0.4* (-0.3) + 0.35 =0.45 (3 marks)

Binary sigmoid (x) = 0 .61064227354 (3 marks)

Q. 3 calculation of xmean and ymean (2+2 marks)

S_no	x	y	x- mean(x)	y-mean(y)
1	95	85	17	8
2	85	95	7	18
3	80	70	2	-7
4	70	65	-8	-12
5	60	70	-18	-7
Sum	390	385		
Mean	78	77		

S_no	x	y	(x- mean(x))^2	(y-mean(y))^2
1	95	85	289	64
2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

S_no	x	y	(x- mean(x)) * (y-mean(y))
1	95	85	136
2	85	95	126
3	80	70	-14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

solve for the regression coefficient (b_1): (4marks)

$$b_1 = \Sigma [(x_i - x)(y_i - y)] / \Sigma [(x_i - x)^2]$$

$$b_1 = 470/730$$

$$b_1 = 0.644$$

Once we know the value of the regression coefficient (b_1), we can solve for the regression slope (b_0): (3 marks)

$$b_0 = y - b_1 * x$$

$$b_0 = 77 - (0.644)(78)$$

$$b_0 = 26.768$$

Therefore, the regression equation is: $\hat{y} = 26.768 + 0.644x$.

y=0.644x+26.78

calculation of SSE (5 marks)

X	Y	predicted y	error	Squared error	
95	85	87.948	2.948	8.690704	
85	95	81.508	-13.492	182.0341	
80	70	78.288	8.288	68.69094	
70	65	71.848	6.848	46.8951	
60	70	65.408	-4.592	21.08646	
			SSE	327.3973	

value of predicted y for x=80 (2.75 marks)

$$\hat{y} = b_0 + b_1x$$

$$\hat{y} = 26.768 + 0.644x = 26.768 + 0.644 * 80$$

$$\hat{y} = 26.768 + 51.52 = 78.28$$

Q. 4 (4 marks for basic calculation below, 3 marks for each iteration (3+3)) (total 10)

x0	x1	x2	y	z	ypred	y-ypred	(y-ypred)*x1	(y-ypred)*x2
1	2.7	2.5	0	1	0.73	-0.73	-1.97	-1.83
1	3	3	0	1	0.73	-0.73	-2.19	-2.19
1	5.9	2.2	1	1	0.73	0.27	1.59	0.59
1	7.7	3.5	1	1	0.73	0.27	2.07	0.94
						-0.23	-0.13	-0.62

Iteration 1

b0	0.93
b1	-0.04
b2	-0.19

x0	x1	x2	y	z	ypred	y-ypred	(y-ypred)*x1	(y-ypred)*x2
1	2.7	2.5	0	3.14	0.96	-0.96	-2.59	2.40
1	3	3	0	0.26	0.56	-0.56	-1.69	1.69
1	5.9	2.2	1	0.29	0.57	0.43	2.52	1.26
1	7.7	3.5	1	-0.02	0.50	0.50	3.88	1.74
						-0.15	0.53	1.77

Iteration 2

b0	0.89
b1	-0.08
b2	0.34

(Calculation of SSE = 6 marks)

x0	x1	x2	y	z	ypred	y-ypred	SSE
1	2.7	2.5	0	1.53	0.82	-0.82	0.67

1	3	3	0	1.67	0.84	-0.84	0.71
1	5.9	2.2	1	1.16	0.76	0.24	0.06
1	7.7	3.5	1	1.46	0.81	0.19	0.04
							1.48

for given scenario of (2.75 marks)

x1 = 6.2 x2 = 3.1

prob 0.8089

therefore class = 1

Q.5 Naïve bayes classifier:

Assumption of conditional independence : presence of a particular feature in a class is unrelated to the presence of any other feature. Features are independent

Advantage is : calaulation of probabilities becomes easy and Naive Bayes classifier performs better than other models with less training data

(total 3 marks)

Calculations :

For (PhD Student, Class=No):

– If Class=No

• sample mean = 110

• sample variance = 2975

For loan to be paid: If class=No: sample mean=110 sample variance=2975

If class=Yes: sample mean=90 sample variance=25 (4.75 marks for mean and variance of both the classes)

P(loan=120 | class= no)= 0.0072

P(Ph.D Student =Yes|No) = 3/7
P(Ph.D Student =No|No) = 4/7
P(Ph.D Student =Yes|Yes) = 0
P(Ph.D Student=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/7
P(Marital Status=Divorced|Yes)=1/7
P(Marital Status=Married|Yes) = 0 (5 marks for calculation of probabilities)

P(X|Class=No) = P(Ph.D Student =No|Class=No) * P(Married| Class=No) * P(loan to be paid=120K| Class=No)
= 4/7 * 4/7 * 0.0072 = 0.0024 (2 marks)

- P(X|Class=Yes) = P(Ph.D Student =No| Class=Yes)* P(Married| Class=Yes) * P(loan to be paid=120K| Class=Yes)
= 1 *0 * 1.2 * 10-9 = 0 (2 marks)

Since P(X|No)P(No) > P(X|Yes)P(Yes)

Therefore P(No|X) > P(Yes|X) => Class = No (2 marks)

Q.6 (Confusion Matrix 4 marks)

10000	Positive	negative
Positive	TP 620	FP 180
Negative	FN 380	TN 8820

One marks each (total 5)

Sensitivity = [tp /(tp+fn)] *100 = 62%
Specificity = [tn/(tn+fp)]+100 = 98%
TPR = tp / (tp+fn)] = 0.62

FPR= fp/(fp+tn) = 0.02

accuracy =
[(tp+tn)/(tp+tn+fp+fn)]*100 = 0.944

MSE function is non-convex for binary classification. Thus, if a binary classification model is trained with MSE Cost function, **it is not guaranteed to minimize the Cost function**. (2.5)

Moreover in case of multiclass classification also less RMSE does not mean more accurate model. (2.5)

Representational power of feed forward networks: (1.5 each , name + brief explanation) total 4.5

a. Boolean function
b. Continuous function
c. Arbitrary function