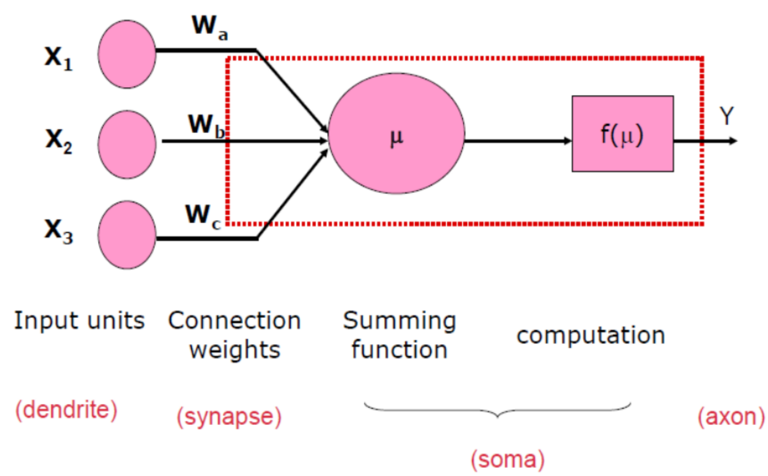
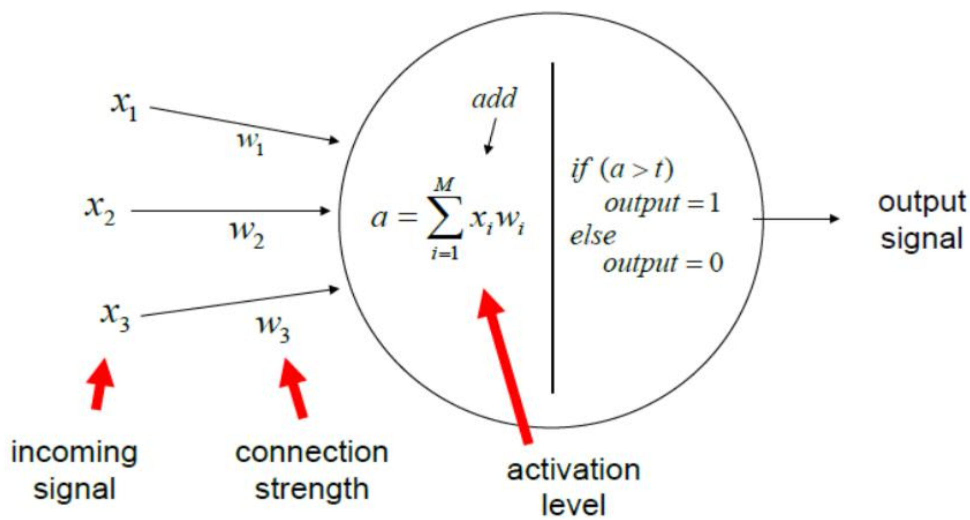
Neural Networks

Recap

Networks of processing units (neurons) with connections (synapses) between them.





Activation Functions

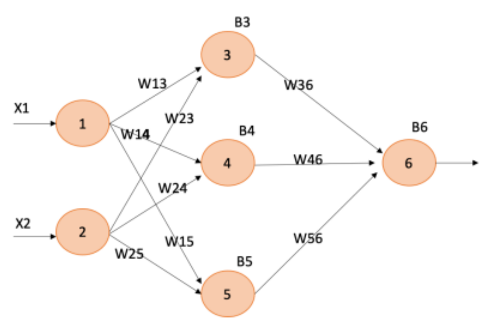
1. Step Function - outputs 0 for any negative input and 1 for any non-negative (zero or positive) input.
2. Linear Function - returns the input value itself without any transformation.
3. ReLu Function - returns 0 for negative inputs and the input value itself for non-negative inputs, effectively introducing non-linearity to the network.
4. Sigmoid Function - maps the input to a value between 0 and 1, which is useful for binary classification problems.

Question

1. You are training a Multilayer Perceptron (MLP) neural network for a particular classification task. After some investigation, your neural network is constructed with 5 input variables, one hidden layer with 12 nodes and one output layer with 3 nodes (the classes). How many network parameters are required to be tuned/trained? Show your detailed calculations. (4 marks)

2.

Consider the below neural network architecture containing 6 neurons.



X1 and X2 are two input features. Wights between corresponding neurons are denoted using W and bias of neurons are mentioned by B values.

The below table shows a part of training data instances.

X1	X2	Y
1	0	0
0	1	0
1	1	1

The following table summarizes the initial weights and biases of neural network.

W13	W23	W14	W24	W15	W25	W36	W46	W56	B3	B4	B5	B6
0.2	- 0.3	0.4	0.1	0.5	- 0.6	- 0.5	0.6	0.3	- 0.3	0.1	0.4	- 0.7

Answer all the questions by showing all the steps of your work.

i) Assuming the Sigmoid is the activation function, execute the forward propagation and calculate the inferred value by the neural network using the first data instance in the training data set. Hence calculate the error associated with the inference task. (12 Marks)

Sigmoid activation function for Z is:

$$f(z) = \frac{1}{1 + e^{-Z}}$$

Error at the output layer is given by the equation:

$Err_j = O_j(1 - O_j)(T_j - O_j)$; For the calculation purpose you can safely assume $T_j = 1$

ii) Using the error value calculated above, execute back propagation and calculate error values in neurons denoted by 3, 4 and 5 as well. (6 Marks)

Error of a hidden layer neuron is given by:

$$Err_j = O_j(1 - O_j)\sum_k w_{j,k} \cdot Err_k$$

i, j, k denotes identifiers of neurons and W, B, Err and O are weights, biases, errors and outputs respectively.

iii) Now considering the Learning Rate (l) is 0.7, calculate the updated weights and bias values resulted by the back propagation. (10 Marks)

Updated weight values are given by:

$$w_{i,j} = w_{i,j} + \Delta w_{i,j} \text{ } where \, \Delta w_{i,j} = l \cdot Err_j \cdot O_i$$

Updated bias values are given by:

$$B_j = B_j + \Delta B_j \text{ } where \text{ } \Delta B_j = l \cdot Err_j$$

I'm showing only the answers here, make sure you follow all the steps in exam. Use at least 3 decimal points

I3=-0.3, I4=0.5, I5=0.9, O3=0.475, O4=0.622, O5=0.711

I6=-0.351 O6=0.413

Err6=0.142

Err3=0.018

Err4=0.020

Err5=0.009

B3=-0.312

B4=0.114

B5=0.406

B6=-0.6

W56=0.37, W46=0.662, W36=-0453, W25=-0.6, W15=0.506, W24=0.1, W14=0.414, W23=-0.3, W13=0.187

Partitioning Clustering

Recap

Grouping together similar data points.

Similarity/Dissimilarity Between Objects

- Distances are normally used to measure the similarity or dissimilarity between two data objects.
    - Euclidean distance: Root of sum of squared differences of distances across dimensions.
    - Manhattan distance: absolute sum of differences of distances across dimensions.

Intra-cluster vs Inter-cluster

- Intra-cluster distances are minimized.
- Inter-cluster distances are maximized.

Question

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

	Individual	Mean Vector
Group 1	1	(1.0, 1.0)
Group 2	4	(5.0, 7.0)

Evaluation Metrics

Classification

Recap

F1 Measure
F1 measure combines Precision and Recall and allows for easier comparison of two or more algorithms.
F1 = 2 * Precision * Recall / (Precision + Recall)

Specificity
Specificity = TN / (TN+FP). The specificity measures how accurate is the classifier in not detecting too many false positives (it measures its negativity)

Sensitivity or Recall
Sensitivity = TP/ (TP + FN). Measures the classifier ability to detect positive classes (its positivity)

Precision
Precision = TP / (TP + FP)

Question

The SoftMax function was employed instead of the sigmoid activation function in the output layer for the model. After retraining the model, a confusion matrix was generated for the three classes based on the maximum probability prediction.

		Predicted Class		
		Positive (P)	Negative (N)	Neutral (O)
Actual Class	Positive (P)	45	6	9
	Negative (N)	2	52	6
	Neutral (O)	7	11	42

1. Find the Overall Accuracy of the confusion matrix (CM) **(1 mark).**
2. Decompose the above given CM into individual CM for **Positive (P)** and **Negative (N)** classes **(4 marks)** and calculate the specificity and sensitivity for each of these classes **(2 marks).**
3. Calculate F1 scores for each class.

Regression

Recap

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error
n = number of data points
Y_i = observed values
$\hat{Y}_i$ = predicted values

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}$$

RMSD = root-mean-square deviation
i = variable i
N = number of non-missing data points
x_i = actual observations time series
$\hat{x}_i$ = estimated time series

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

MAE = mean absolute error
y_i = prediction
x_i = true value
n = total number of data points

Question

Period	Expected	Actual	Difference	Square of Differences
Jan 2015 - Jun 2015	166	160	-6	36
Jul 2015 - Dec 2015	170	172	2	4
Jan 2016 - Jun 2016	180	176	-4	16
Jul 2016 - Dec 2016	186	186	0	0
Jan 2017 - Jun 2017	194	196	2	4
Jul 2017 - Dec 2017	208	202	-6	36
Jan 2018 - Jun 2018	208	212	4	16

MSE =	16.0000
RMSE =	4.0000