Chapter 7

1. What are the basics of the neuron's model?

The first part of a neuron is the set of inputs xi, also called synapses, each of these inputs have a weight where k is the neuron it is being input too and i is the input synapse. After the input is put into the neruon with the coresponding wieght an adder summes th input signals in a linear combiner for the output of the neruon. The output of the neroun is desided from this sumation by using it in an activation function that keeps the output limited at a certain ampllitude. There might also be a bias used in neroun that increases or losers the net input of the activation function making a neroun produce an output more often.

1. Discuss the differences between feedforward and recurrent neural networks.

A feedforward network processes from the input side to the output side with no loops or propagation, meaning that it will run through the network going layer to layer until it creates an output with no nodes in the same layer connecting to each other. This makes the network modular where each node has the same functionality when creating the next output to feed to the next node.

A recurrent neural network is one that loops. This looping is usually with a delay element to synchronize the components but the process is defined by how each output will loop back to the node the created the output to process again.

1. For the given competitive network:
   1. Find the output vector [Y1, Y2, Y3] if the input sample is [X1, X2, X3] = [1, -1, -1]

Net(1)` = 0.7\*x1 + -0.3\*x3 = 0.7\*1 + -0.3\*-1 = 1

Net(2)` = 0.3\*x1 + -0.7\*x2 = 0.3\*1 + -0.7\*-1 = 1

Net(3)` = 0.3\*x2 + 0.7\*x3 = 0.3\*-1 + 0.7\*-1 = -1

Net(1) = net(1)` + 0.5\*1 + 0.5\*-1 = 1 MAX

Net(2) = net(2)` + 0.3\*1 + 0.6\*-1 = 0.7

Net(3) = net(3)` + 0.2\*1 + -0.1\*1 = -0.9

Competition between outputs shows that net(1) is the highest so the output should be

Y = {1, 0, 0}

1. What are the new weight factors in the network?

Learning rate of 0.2

W11 = 0.7 + 0.2\*(0-0.7) = 0.56

W21 = 0 + 0.2\*(1-0) = 0.2

W31 = -0.3 + 0.2\*(1 - -0.3) = -0.04

1. Which type of artificial neural-network architecture does not contain a hidden layer? Why?

a) Backpropagation

b) Perceptron

Has no hidden layers. A perceptron network that is not multilayered is just a single layer network. It works as a simple feedforward network differently from its MLP (Multilayer Perceptron) counterpart.

c) Self-organizing map

d) Convolutional networks

e) Several previous types

1. What Is the Difference Between Epoch, Batch, and Iteration in Deep Learning?

Epoch is how many times the network will learn from the current data set. Meaning that for each epoch that is the number of times it has run the training set through the network and learned from it.

Batch size is used to refer to how many samples are used when updating a network. It can either be size of the training data as a whole or the size of some of the training data. Either way the batch is essentially the data that is used during training.

An iteration is the number of times network will have to run to train on all of the training set. Say that the batch sizes are only half of the training data then the number of iterations to get through both batches and all of the training data will be 2.

1. What Is Overfitting and Underfitting, and How to Combat Them?

Underfitting happens when the model is trained on to few samples, this makes it difficult to actually learn from the data set it uses. To combat this more samples are used to give the model information to learn from. This can also be done by increasing the number of features from the current samples that are used.

Overfitting happens when the model is learning to focus on features only in the training set and not the general features of the data. To fix this you can use more samples like with underfitting to try and introduce more information to that it can learn general features from, or you can do some techniques like regularization to make the data less feature heavy or just removing features entirely.

Chapter 8

1. Explain the basic idea of the ensemble learning, and discuss why the ensemble mechanism is able to improve prediction accuracy of a model.

Ensemble learning is the idea of using multiple predictive models trained on the same training data to lower the error rate of prediction. The process is easy to understand but requires a little thought to understand how it can make predictions more accurate. The idea is that since each of these models presumably has a low error rate than the ensemble model that is based on the outputted predictions of all the other models will be less likely to miss predict. In an example seen in the text if 15 models are used with each having an error rate of 0.3 then the ensemble model will only miss predict when over half the models also miss predict. This makes the error rate of the overall model much lower then 0.3 and from the math in the book, in this example the error rate of the ensemble model is 0.05. This is the case because it is less likely for a majority of the models to predict incorrectly since they have low error rates.

Chapter 9

1. Given a set of five dimensional categorical samples:

A=(1, 0, 1, 1, 0)

B=(1, 1, 0, 1, 0)

C=(0, 0, 1, 1, 0)

D=(0,1, 0, 1, 0)

E=(1, 0, 1, 0, 1)

F=(0, 1, 1, 0, 0)

Apply agglomerative hierarchical clustering using

Single line similarity measure based on Rao's coefficient.

A=(1, 0, 1, 1, 0)

B=(1, 1, 0, 1, 0)

a = where A and B are 1, b = A is 1 and B is 0, c = A is 0 and B is 1, d = A and B are 0

a = 2, b= 1, c = 1, d=1

RC(A, B) = 2/(2+1+1+1) = 0.4

A=(1, 0, 1, 1, 0)

C=(0, 0, 1, 1, 0)

a=2, b=1, c=0, d=2

RC(A, C) = 2/(2+1+0+2) = 0.4

RC(A,D) = 1/(1+2+1+1) = 0.2

RC(A,E) = 2/(2+1+1+1) = 0.4

RC(A,F) = 1/(1+2+1+1) = 0.2

RC(B,C) = 1/(1+2+1+1) = 0.2

RC(B,D) = 2/(2+1+0+2) = 0.4

RC(B,E) = 1/(1+2+2+0) = 0.2

RC(B,F) = 1/(1+2+1+1) = 0.2

RC(C,D) = 1/(1+1+1+2) = 0.2

RC(C,E) = 1/(1+1+2+1) = 0.2

RC(C,F) = 1/(1+1+1+2) = 0.2

RC(D,E) = 0/(0+2+3+0) = 0

RC(D,F) = 1/(1+1+1+2) = 0.2

RC(E,F) = 1/(1+2+1+1) = 0.2

First cluster (D, E) since it has the smallest value so the cluster is (D,E) but then we move to the next smallest value found which is 0.2 so all values with 0.2 that are not in cluster are added. These are (A,D),

(A,F), (B, C), (B, E), (B, F), (C, D), (C, E), (C, F), (D, F), and (E, F). This makes the final and only cluster

(D, E, F, C, B, A)

Complete link similarity measure based on simple matching coefficient SMC.

SMC(X, Y) =

SMC(A, B) = (2+1)/(2+1+1+1) = 0.6

SMC(A, C) = (2+2)/(2+1+0+2) = 0.8

SMC(A,D) = (1+1)/(1+2+1+1) = 0.4

SMC(A,E) = (2+1)/(2+1+1+1) = 0.6

SMC(A,F) = (1+1)/(1+2+1+1) = 0.4

SMC(B,C) = (1+1)/(1+2+1+1) = 0.4

SMC(B,D) = (2+2)/(2+1+0+2) = 0.8

SMC(B,E) = (1+0)/(1+2+2+0) = 0.2

SMC(B,F) = (1+1)/(1+2+1+1) = 0.4

SMC(C,D) = (1+2)/(1+1+1+2) = 0.6

SMC(C,E) = (1+1)/(1+1+2+1) = 0.4

SMC(C,F) = (1+2)/(1+1+1+2) = 0.6

SMC(D,E) = (0+0)/(0+2+3+0) = 0

SMC(D,F) = (1+2)/(1+1+1+2) = 0.6

SMC(E,F) = (1+1)/(1+2+1+1) = 0.4

First merges on (D,E) with 0 value. Then it merges (B,C) and (A,F) on the values of 0.4 creating two more clusters which gives three clusters. (B,C) and (A,F) are then merged on the values of 0.6 and finally (D,E) is merged on the values of 0.8.

1. Plot the dendrograms for the solutions to part (i) and (ii) of (a).

Diagram

Description automatically generated

Diagram

Description automatically generated

1. Given the similarity matrix between five samples:
   1. Use the similarity matrix in the Table to perform complete-link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | P5 |
| P1 | 1 |  |  |  |  |
| P2 | .1 | 1 |  |  |  |
| P3 | .41 | .64 | 1 |  |  |
| P4 | .55 | .47 | .44 | 1 |  |
| P5 | .35 | .98 | .85 | .76 | 1 |

Merge on P1 and P2 take min for other values in the same column

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | P1P2 | P3 | P4 | P5 |
| P1P2 | 1 |  |  |  |
| P3 | .41 | 1 |  |  |
| P4 | .47 | .44 | 1 |  |
| P5 | .35 | .85 | .76 | 1 |

Merge on P5

|  |  |  |  |
| --- | --- | --- | --- |
|  | P1P2P5 | P3 | P4 |
| P1P2P5 | 1 |  |  |
| P3 | .41 | 1 |  |
| P4 | .47 | .44 | 1 |

Merge on P3

|  |  |  |
| --- | --- | --- |
|  | P1P2P5P3 | P4 |
| P1P2P5P3 | 1 |  |
| P4 | .44 | 1 |

Merge on P4

Chart, diagram, line chart

Description automatically generated

If we take the max values instead we get a dendrogram that looks like

Diagram

Description automatically generated

1. How many clusters exist if the threshold similarity value is 0.5. Give the elements of each cluster.

(P1,P2) = .1

(P1,P3) = .41

(P1,P4) = .55

(P1,P5) = .35

(P2,P3) = .64

(P2,P4) = .47

(P2,P5) = .98

(P3,P4) = .44

(P3,P5) = .85

(P4,P5) = .76

All values that have a distance of less then 0.5 will be able to be a valid cluster so that makes the highlighted point pairs above a valid cluster that can be made. So the total number of clusters is 5.

1. If DBSCAN algorithm is applied with threshold similarity of 0.6 and MinPts >= 2 (required density) what are core, border, and noise points in the set of points pi given in the table. Explain

A point is a core point if it has more than min pts within neighborhood. P1 has 4 points that all fit with in the similarity threshold so they would be included in its neighborhood making it a core point. This along with P4 which has three points that are less than the threshold similarity and, in its neighborhood, makes it also a core point. Core Points: P1 and P4

A boarder point is one that has fewer than the min points in its neighborhood and is in the neighborhood of a core point. This would apply to all the other points that are not core, P2, P3, and P5, because they are all in the neighborhood of P1 since their distance is all under 0.6 from P1. Boarder Points: P2, P3, and P5

A noise point is any that is not a core or a border point. So, there are no noise points.

1. Answer True/False to the following statements. Discuss your answer if necessary.
   1. Running K-means with different initial seeds is likely to produce different results

True: a different seed would change the randomness of it

1. Initial cluster centers have to be data points

True: They contain data points in them, so they have to be data points

1. Clustering stops when cluster centers are moved to the mean of clusters.

Ture: it is designed to find the mean of clusters so this would be its stopping point

1. K-means can be less sensitive to outliers is standard deviation is used instead of the average.

False: it is sensitive to outliers

1. K-means can be less sensitive to outliers if median is used instead of the average.

True: if it is not the average it becomes less sensitive