Foundation of ML Quiz - 3

Duration of Quiz is 30 minutes (strict), 9 AM to 9:30 AM. You can resubmit until 9:30 AM. No negative marks, but questions carry unequal marks.

Program (PhD, MTech, MDS, BTech) *

- ⦿ BTech
- ◯ MTech
- ◯ MDS
- ◯ PhD

Name *

A Venkata Sai Mahesh

Roll no *

CS18BTECH11001

Course ID (FoML course ID you have registered in AIMS : CS5590, AI5000, SM5000, AI2000) *

○ CS5590 ●

○ AI5000

○ SM5000

○ AI2000

The single link agglomerative clustering algorithm groups two clusters using minimum distance between points in the two clusters

● True

○ False

Consider minimize f(x, y) = x + 2y, subject to x^2 + y^2 – 4 = 0. When you use lagrange multiplier based approach to solve this constrained optimisation problem, what will be the value of lagrange multiplier?

○ +/- 2/sqrt(5)

○ +/- 4/sqrt(5)

○ 2/sqrt(5)

● +/- sqrt(5)/4

A multiple-layer neural network with linear activation functions is equivalent to one single-layer perceptron that uses the same error function on the output layer and has the same number of inputs

- ◉ True
- ○ False

Q13. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithm for the same dataset?

- ○ Proximity funtion used
- ○ data points used
- ○ Variables used
- ◉ All of the above

A single layer neural network with a sigmoid activation for binary classification with the cross entropy loss is exactly equivalent to logistic regression.

- ◉ True
- ○ False

Consider clustering of documents which are represented in bag of words form (Each dimension correspond to a word in vocabulary, and feature value corresponds to number of times word has appeared in the document). Which of the following is most appropriate ?

○ Student' t mixture model

● Multinomial mixture model

○ Gaussian mixture model

○ Bernoulli mixture model

A 3-input neuron is trained to output a zero when the input is 110 and a one when the input is 111. After generalization, the output will be zero when and only when the input is:

○ 000 or 110 or 011 or 101

○ 010 or 100 or 110 or 101

● 000 or 010 or 110 or 100

○ 100 or 111 or 101 or 001

Using the kernel trick, one can get non-linear decision boundaries using algorithms designed originally for linear models
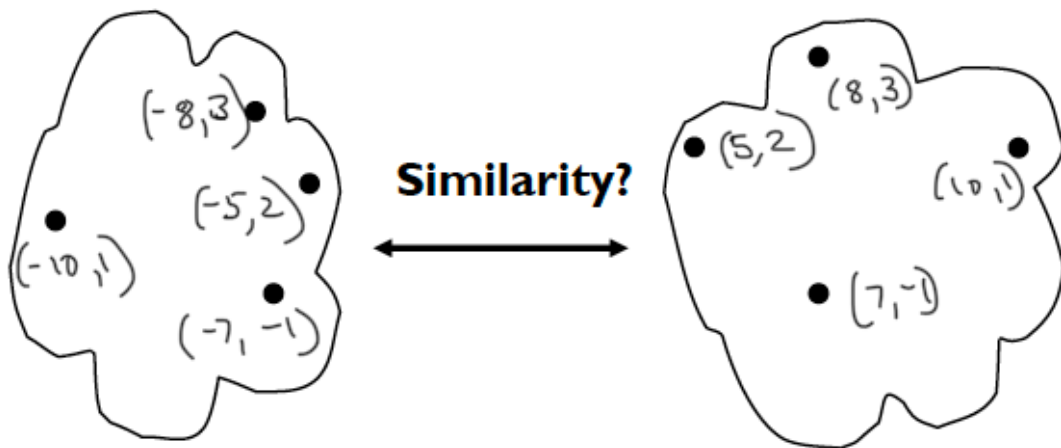
● True

○ False

You use a multilayer neural network and notice that the training error is going down and converges to a local minimum. When you test on the new data, the test error is abnormally high. Which among the following would you not recommend ?

○ The training data size is not large enough. Collect a larger training data and retrain it.

○ Use the same training data but add two more hidden layers.

○ Play with learning rate and add regularization term to the objective function.

◉ Use a different initialization and train the net-work several times.

○ Option 5

Consider two clusters and data points in these two clusters in the Figure below. If you use centroid based agglomerative clustering what will be the cluster similarity ?

○ 100

◉ 225

○ 256

○ 400

○ 238

Which of the step is not required for K-means clustering?

○ a distance metric

○ initial number of clusters

○ initial guess of clusters centroids

⦿ None

[True/False] If two classes are linearly inseparable, perceptron can convergence

○ True

⦿ False

Consider two clusters and data points in these two clusters in the Figure below. If you use complete-link agglomerative clustering, what will be the cluster similarity ?
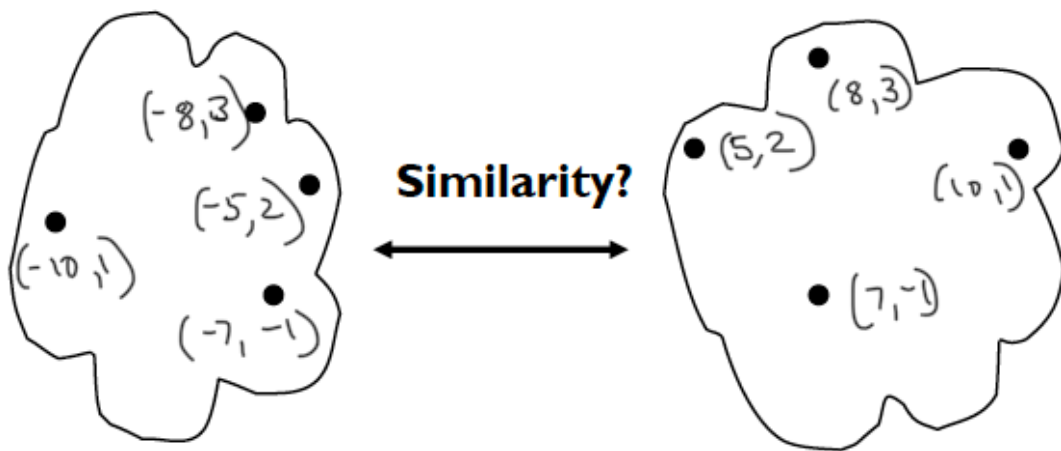
○ 100

○ 200

◉ 400

○ 256

○ 238

What are some practical problems with the sigmoidal activation function in neural nets?

○ It is convex, and convex functions cannot solve nonconvex problems

◉ It does not work well with the entropy loss function

○ It can have negative values

○ it can lead to the vanishing gradient problem

The numerical output of a tanh node in a neural network :

◉ Is bounded between -1 and 1.

○ Is bounded between 0 and 1.

○ is unbounded

○ is unbounded, encompassing all positive real numbers

Your neural network consists of an input layer with 40 units, a hidden layer with 20 units, and an output layer with 2 units. Use the ReLU activation function for the hidden units and no activation function for the outputs. The number of parameters (weights) in this network considering bias term is

○ 862

◉ 840

○ 1600

○ 924

Which of the following can help to reduce overfitting in an SVM classier?

◉ Use of slack variables

○ Normalizing the data

○ High-degree polynomial features

○ Setting a very low learning rate

k-means algorithm does coordinate descent on a non-convex objective function

◉ True

○ False

If K1 and K2 are kernels, K1*K2 (product) is a kernel.

◉ True

○ False

Consider a perceptron classifier for 4 dimensional data with weights [0.2, -0.8, 0, 0.5]. The inputs x1 = [1,0,1,1] and x2 = [1,1,1,1] will be classified respectively to the classes

○ -1 and -1

◉ +1 and +1

○ +1 and -1

○ -1 and +1

Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means

◉ Yes

○ No

The loss function in a multi-layer neural network is convex with respect to weight vectors.

○ True

◉ False

The numerical output of a ReLU node in a neural network :

○ Is bounded between 0 and 1.

○ Is bounded between -1 and 1.

○ is unbounded, encompassing all positive real numbers

◉ is unbounded

Which version of clustering algorithm is more sensitive to outliers

◉ K means

○ K modes

○ K medians

○ K- Medoid

You want to cluster this data into 2 clusters. Which of the these algorithms would be most appropriate ?

- ○ K-Mean
- ○ Single linkage clustering
- ⦿ Gaussian mixture model
- ○ Average linkage clustering

You are increasing the size of the layers (more hidden units per layer) in your neural network. What kind of impact it will have on bias and variance?

- ○ a) increases, increases
- ○ b) increases, decreases
- ⦿ c) decreases, increases
- ○ d) decreases, decreases.

Suppose you are given an EM algorithm that finds maximum likelihood estimates for a model with latent variables. You are asked to modify the algorithm so that it finds MAP estimates instead. Which step or steps do you need to modify?

- ○ Expectation
- ● Maximization
- ○ No modification necessary
- ○ Both

In perceptron learning, what happens when input vector is correctly classified?

- ○ large adjustments in weight is done
- ● no adjustments in weight is done
- ○ small adjustments in weight is done
- ○ weight adjustments doesn't depend on classification of input vector

Which of the following is/are true regarding an SVM?

- ○ For two dimensional data points, the separating hyperplane learnt by a linear SVM will be a straight line.
- ● In theory, a Gaussian kernel SVM cannot model any complex separating hyperplane.
- ○ For every kernel function used in a SVM, one can obtain an equivalent closed form basis expansion.
- ○ Overfitting in an SVM is not a function of number of support vectors

You are given the dataset: A = (1,0), B = (0,1) and C = (2,0). The k-means algorithm is initialised with centres at B and C. Upon convergence, the centroids will be at

○ B and C

○ A and the midpoint of BC

◉ C and the midpoint of AB

○ B and the midpoint of AC

In kernel support vector machines, the Gaussian kernel

○ is less prone to oscillating than polynomials.

◉ lead to Gaussian shaped decision boundaries

○ has good properties in theory but is rarely used in practice

○ is equivalent to mapping to a space whose dimension is exponential in original dimension.

Average link agglomerative clustering can easily cluster different sized and non convex shaped clusters.

◉ True

○ False

You are given a labeled binary classification data set with N data points and D features. Suppose that N < D. In training an SVM on this data set, which of the following kernels is likely to be most appropriate?

◉ Linear kernel

◯ Quadratic kernel

◯ Higher-order polynomial kernel

◯ RBF kernel

Consider a neuron which uses the logistic function l(x) for non linear transformation. The gradient of the logistic function is

◯ 1-l(x)

◯ l(x)(l(x)-1)

◯ l(x)(1+l(x))

◉ l(x)(1-l(x))