COSC 3337

Review on Nov. 9, 2022

For the Nov.14 Midterm2 Exam

Also take a closer look at the recent GHC presentations which gave a review of other material which is relevant for Midterm2.

**1) Neural Networks**

a) Describe how multi-layer neural networks, consisting of 3+ layers learn a model for a training set! Limit you answer to at most 9 sentences! [7]

Neural network learning tries to find weights that minimize the error in the neural network prediction for a training set [1]. Neural networks employ gradient decent hill climbing to find the “best” weights. [1]. In particular, Neural network learning adjust weights using the gradient of the error function of the training set [1]; the search starts with a random initial weight vector and weights are adjusted in the direction of the steepest negative gradient of this error function---that is weights are updated accordingly moving in the direction that reduces the error the most [2]; The step width of this weight update depends on the learning rate and other factors [1]. In order to apply this procedure, the error for each none-input node has to be known. As this error is not initially given intermediate for intermediate layer nodes, it is computed using the back-propagation algorithm [2].

Other observation might deserve credit. At most 7 points!

b) Assume the following subset of a neural network is given:

wA,B=0.2

ΔB=0.4

A B

Assume that B is an intermediate node of a neural network, the forward propagation activation values of nodes aA and aB are 0.5 and 0.8 and the current value of WA,B is 0.2; the associated error ΔB of node B that was computed by the back propagation algorithm is 0.4, the learning rate γ is assumed to be 0.5. First give the general weight update formula and then compute the new value of weight wA,B! [5]

Formula: wAB= wAB + γ\*aA\*ΔB [1]

where aA denotes the activation of node A; we receive as the new weight of WAB:

WAB=0.2 + 0.5\*0.5\*0.4=0.2+0.1=0.3 No Partial Credit.

c) Looking at the sub neural network in the figure below; what does the back-propagated associated error ΔA for a node A depend on? Give a formula to compute its value, assuming g is the activation function of the neural network.

wA,B=0.5

ΔB=0.4

A B

wC**,A**=0.6 wD,A=0.2

C D

ΔA depends on: the associated error ΔB in the node B, the weight of the connection between A and B and the derivative g’ of the activation function g for the linear input of node A zA.

Formula: ΔA=g’(za)\*wAB\*ΔB=g’(zA)\*0.5\*0.4=g’(0.12+0.08)\*0.5\*0.4=g’(0.2)\*0.5\*0.4

**2. K-Means and K-Medoids/PAM and Clustering in General**

1. Assume we apply K-medoids for k=2 to a dataset consisting of 4 objects numbered 1,..,4 with the following distance matrix:

0 6 5 2 🡨object1

0 4 3

0 1

0 (e.g. the distance between object 2 and 4 is 3)

The current set of representatives is {3,4} (objects 3 and 4); indicate all computations k-medoids (PAM) performs in its next iteration! Does k-medoids get a new set of representatives or does it terminate in the next iteration? [6]

RS={3,4} clusters: {3} (1,2,4} SEE=2\*\*2+3\*\*2

New Represnetative sets are created

{1,4} …. SSE=3\*\*2+1\*\*2

{2,4} {2} {1,3,4} SSE=2\*\*2+1\*\*2

{1,3} …SSE=4\*\*2+1\*\*2

{2,3} … SSE=5\*\*2+1\*\*2

The SSE decreased and therefore PAN will run for another generation for the “new” representative set {2,4}

One error: at most 3.5 points; 2 errors at most 1 point.

b) Assume the following dataset is given: (1,1), (2,2) (4,4), (5,5), (4,6), (6,4) . K-Means is used with k=2 to cluster the dataset. Moreover, Manhattan distance is used as the distance function (formula below) to compute distances between centroids and objects in the dataset. Moreover, K-Means’s initial clusters C1 and C2 as follows:

C1: {(1,1), (3,3), (4,4), (6,6)}

C2: {(6,4), (4,6)}

Now K-means is run for a single iteration; what are the new clusters you obtain[[1]](#footnote-1) [4]

**d((x1,x2),(x1’,x2’))= |x1-x1’| + |x2-x2’| Manhattan Distance**

centroid C1= (3.5,3.5}

centroid C2= (5,5)

New Clusters

C1={(1,1), (3,3), (4,4)}

C2={(6,6},(4,6), (6,4)}

One error at most 1.5 points; 2 errors: 0 points

Problem 2 continued

c) Compare k-means with Hierarchical clustering; what are the main differences in the way they are forming clusters and in general? [4]

K-Means creates a single clustering and HC creates a hierarchy of object sets; that is, multiple clusterings [2]

HC creates a dendrogram by merging the closest clusters[1]; K-means creates clusters by assigning the objects in a dataset to the closest centroid [1]

**3. Parametric Density Estimation**

[GroupF-PDE-Task.pptx (live.com)](https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww2.cs.uh.edu%2F~ceick%2FUDM%2FGroupF-PDE-Task.pptx&wdOrigin=BROWSELINK)

**4. Non-Parametic Density Estimation**

Discuss new slide added to: [NPDE.pptx (live.com)](https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww2.cs.uh.edu%2F~ceick%2FUDM%2FNPDE.pptx&wdOrigin=BROWSELINK)

Also take a look at the solution sketches of Midterm1 NPDE problem.

**5. Miscellaneous Questions**

a. What is an outlier in DBSCAN?

An outlier a non-core point that does not lie within an ε-radius of a core point[[2]](#footnote-2).

b. What is the purpose of validation sets when learning classification and prediction models? What is the purpose of test sets when learning classification and prediction models? [2]

validation set: used to select hyper parameter during training

test set: used to assess the accuracy for classifies unseen examples in straining or used to estimate the generalization error.

c) An unnamed scientist uses training and testsets for supervised learning that share 25% of the examples—25% of the examples in the training set belong to the corresponding test set. So you believe this is a good idea? Give reasons for your answer! [3]

No [1] The purpose of testsets is to estimate the generalization error/to assess how well the classifier work of unseen examples. As the approach uses examples used in training in the testset, the obtained test set accuracies are no longer a good estimate for the generalization error, as examples used in training are used to assess the testset accuracy. Comment: Using training set examples in test sets is even considered as cheating by the ML community.

1. If there are any ties, break them whatever way you want! [↑](#footnote-ref-1)
2. If a non-core point lies within the ε-radius of a core point it is a border point! [↑](#footnote-ref-2)