Christoph F. Eick

**Review COSC 6335 Fall 2020 Final Exam**

December 3, 2020

1) Neural Networks

###  Which of the following statements about neural networks is not true (3 points)

### a. Backpropagation is not needed for 2-layer neural networks

### b. NN learning employs gradient descent search

### c. Increasing the learning rate increases the stepwidth used in gradient descent search

### d. Using a low learning rate speeds up neural network learning

Which of the following statements about neural networks is not true? (3 points)

a. Backpropagation is not needed for neural networks with fewer than 4 layers.

b. NN learning uses the gradient of the trainset error function to determine the direction of the search

c. Decreasing the learning rate decreases the step width used in gradient descent search

d.Using a lower learning rate slows down neural network learning

Correct Answers: d then a

**2) Clustering [17]**

1. What objective function does K-means minimize[[1]](#footnote-1)? [2]

**The sum of the squared distance of the objects in the dataset to the centroid of the cluster they are assigned to**

1. When does K-means terminate? When does PAM/K-medoids terminate? [2]

**When the clustering does not change; when there is no improvement with respect the objective function PAM minimizes with respect to the (n-k)\*k newly generated clusterings.**

1. Assume K-Means is used with k=3 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’ initial clusters C1, C2, and C3 are as follows:

C1: {(2,2), (6,6)}

C2: {(4,6), (8,0)}

C3: {(4,8), (6, 8)}

}

Now K-means is run for a single iteration; what are the new clusters and what are their centroids? [3]

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

**C1 centroid: (4,4) {(2,2), (4,6)} new centroid: (3,4)**

**C2 centroid: (6,3) {(6,6), (8,0)} new centroid: (7,3)**

**C3 centroid: (5,8) {(4,8), (6,8)} centroid: (5,8)**

**Remark: Assigning (6,6) to cluster C3 instead, is also correct!**

d) The DENCLUE algorithm uses density functions to form clusters. How are density functions created by the DENCLUE algorithm from datasets? What are density attractors? What role do density attractors play when forming clusters? [4]

The density for a query point is computed by summing up the influences of the points in the dataset to the query point⎯the influence of a point to the query point decreases as the points distance to the query point increases. Density attractors are local maxima of the density function. Points that are associated with the same density attractor belong to the same cluster⎯hill climbing is used to find this association.

e) Compute the Silhouette for the following clustering that consists of 2 clusters:

{(0,0), (0.1), (1,1)}, {(1,2), (4,4)}; use Manhattan distance for distance computations. Compute each point’s silhouette; interpret the results (what do they say about the clustering of the 5 points; the overall clustering?)![6]

(5.5-1.5)/5.5

(4.5-1)/4.5

(3.5-1.5)/3.5

(2-5)/5

(5-7)/7

In general, the silhouette of the first 3 points is good, the silhouette of the 4th point is bad, because this point has been associated with the wrong clustering, and the silhouette for the 5th points is mediocre because the inter-cluster distance is high due to the incorrect assignment of the point (1.2). The quality of the first cluster is decent, whereas the quality of the second cluster and the overall clustering is poor!

f) Assume you apply k-means to a dataset which contains outliers; assume you apply k-means to a 2D-dataset Y={(-100,-100), (0,0), (1,1), (0, 1), (1, 2), (5, 4), (5,5) , (5,6)} with k=2; how do outliers (e.g. the point (-100,-100)) impact the k-means clustering result? Propose an approach that alleviates the strong influence of outliers on clustering results. If your approach would be used to cluster dataset Y; how would the result change? [6]

Leads to a clustering where the outlier forms a single cluster [2]

Method [2.5]

After method is applied points 2-5 form a cluster, and the last 3 [1.5]

g) What advantages[[2]](#footnote-2) you see in using DBSCAN over K-means? [3]

* Not sensitive to outliers[0.5]; supports outlier detection [1]
* Can detect clusters of arbitrary shape and is not limited to convex polygons. [1.5]
* Not sensitive to initialization [0.5}
* Not sensitive not noise [0.5]

***At most 3 points!!***

h) What cluster models does EM create when called for k=3?

EM creates a model that consists of 3 cluster models, each consisting of:

i. a cluster prior (the 3 priors have to add up to one)

ii. a cluster mean

iii. a cluster covariance matrix

These 9 things can then be used to define a Gaussian Mixture density function; see: <http://pypr.sourceforge.net/mog.html>

3) Spatial Data Mining [5]

What are the main challenges in mining spatial datasets⎯how does mining spatial datasets differ from mining business datasets?

Autocorrelation/Tobler’s first law[1.5], Attribute Space is continuous[0.5], no clearly defined transactions [0.5], complex spatial datatypes such as polygons [1], separation between spatial and non-spatial attributes [1], importance of putting results on maps [0.5], regional knowledge[0.5], a lot of objects/a lot of patterns[0.5].

At most 5 points!

**4) Sequence Mining**

a) Assume the Apriori-style sequence mining algorithm described on pages 429-435 of our textbook is used and the algorithm generated 3-sequences listed below:

Frequent 3-sequences Candidate Generation Candidates that survived pruning

<(1) (2) (3)> (1)(2)(3)(4) (1) (2) (3) (4)

<(1 (2 3)> (1) (2 3) (4) (1) (2 3) (4)

<(1 2 4)> (1) (2 3) (5)

<(1) (2) (4)> (1 2) (3) (4)

<(1) (3) (4)>

<(1 2) (3)>

<(2 3) (5)

<(2 3) (4)>

<(2) (3) (4)>

What candidate 4-sequences are generated from this 3-sequence set? Which of the generated 4-sequences survive the pruning step? Use format of Figure 7.6 in the textbook on page 435 to describe your answer! [5]

5) Outlier Detection

a) Give a brief description of how model-based approaches for outlier detection work.

Fit a statistical model M to the data points of the dataset O; next, the density function dM of the model M is used to assess the likelihood of objects o belonging to O; objects with very values for dM(o) or log(dM(o)) are considered to be outliers in O

b) How do k-nearest neighbor-based outlier detection techniques determine the degree to which “*an object in a dataset is believed to be an outlier*”.

For each object the k-nearest neighbor distance—k is a parameter of the method;—to the other objects in the dataset is computed; objects with very high values for that distance are considered to be outliers

Remark: For example, boxplot based outlier detection approaches could be used to decide which low density/high k-NN distance objects are considered to be outliers for the two approaches, we just described.

**Not covered in this review but relevant for the final exam: fit a parametric or non-parametric model to a dataset---check online credit tasks; design distance function for a dataset; capability to answer “deeper” questions about the DENCLUE 2.0 paper; capability to answer some “more basic” questions about the Autoencoder, CNN, and Spatial Analysis articles (see Review List).**

1. Be clear! [↑](#footnote-ref-1)
2. We are only interested in the advantages and not the disadvantages! [↑](#footnote-ref-2)