URL
Spring Term 2018

Unsupervised Learning Questionnaire

Due date 20/04/2018
Artificial Intelligence Master

Name: Josep Famadas Alsamora 47279799

Instructions

Make sure that you are answering the correct questionnaire, it has to correspond with the identification number you gave when you enrolled the course.

You have to upload the printed file with your answers to the Raco in the entry corresponding to the questionnaire before **april 20th at 23:59**.

You have to open this file using **Acrobat Reader** in order to fill the questionnaire. To begin to answer the questionaire, first click on the **Begin Quiz** text. After you have answered all the question click on the **End Quiz** text. In order to save your answers you must **print** this file as a PDF file, otherwise the answers will be lost when you close the file.

Alternativelly, you can edit the PDF file using your favourite PDF editor, but make sure that the answers you mark are visible.

Begin Quiz

This questionaire is about the different topics from the unsupervised learning part of the course. Each question has a value of 1 point, each question has exactly two correct answers, each incorrect answer discounts 0.5 points of the question.

1. Thinking about dimensionality reduction (multiple answer):

(a) Sparse PCA is a kernelized version of PCA that performs a non linear transformation of the data.
(b) The methods for multidimensional scaling can decide the number of dimensions to keep by using the weights of the combinations of the original attributes.
(c) ✔ Principal Component Analysis is a dimensionality reduction technique that transforms the data so all components are uncorrelated preserving the variance.
(d) ✔ Wrapping is an attribute selection method that uses a model for evaluating the relevance of subsets of the attributes of a dataset.

2. Thinking about dimensionality reduction (multiple answer):

(a) Multidimensional scaling transforms a dataset to a new space with the same dimensionality.
(b) ✔ ISOMAP is a nonlinear multidimensional scaling method that uses the geodesic distance among examples instead of the euclidean distance.
(c) ICA is applied to gaussian data and PCA is applied to non gaussian data.
(d) ✔ Local Multidimensional Scaling tries to preserve in the new space the locality of closer neighbors and to increase the distance of the non neighbors.

3. Thinking about dimensionality reduction (multiple answer):

(a) Least Squares Multidimensional Scaling performs a non linear transformation by minimizing the variance of the data.
(b) ✔ The eigendecomposition obtained using Kernel PCA, unlike PCA, does not give the importance/weights of the components.
(c) ✔ Random Projection performs a linear transformation of the data by generating a random transformation matrix of almost orthogonal vectors.
(d) The Laplacian score is a wrapper attribute selection method based on the K-means algorithm.

4. Thinking about dimensionality reduction (multiple answer):

(a) ✔ The eigenvalues of the decomposition that computes the PCA components are proportional to que variance of the data that they explain.
(b) Local Linear Embedding is a linear multidimensional scaling method because uses the linear reconstruction of the examples by its neighbors as optimization criteria.

(c) Filtering is an attribute selection method that uses a relevance measure that assesses the importance of each attribute individually.
(d) Non linearly separable data can be separate by using a linear kernel with Kernel PCA.

5. Thinking about dimensionality reduction (multiple answer):

(a) Unsupervised Attribute selection methods transform the attributes of the data to a new space with less dimensions.
(b) Multidimensional scaling transforms a dataset by preserving the statistical distribution model of the data.
(c) Nonegative Matrix Factorization obtains a linear transformation where all the weights of the transformation are positive.
(d) The transformation performed by ISOMAP preserves the local neigborhood of the examples in the new space by using the geodesic distance in the transformation.

6. Thinking about processing of missing values (multiple answer):

(a) Using the mean/mode of the k-nearest neighbors of an example is a more accurate method than using the global mean, but it is computationaly expensive.
(b) Examples with missing values will decrease the quality of the discovery process, the more missing information, the more impact.
(c) Discarding the examples with missing values does not change the overall data distribution.
(d) A method for imputing missing values is to learn a model for the attribute with the rest of the attributes and it is not more expensive than computing the mean/mode of the data.

7. Thinking about the detection/processing of outliers (multiple answer):

(a) Detecting abnormalities on the distribution of the distances among the examples is a non parametric method for finding outliers in a dataset.
(b) Kernel Density Estimation is a parametric method for finding outliers in a dataset.
(c) An outlier is defined as an example that does not have a value for some of its attributes.
(d) Computing an histogram of the attributes of the data is a non parametric method for detecting outliers.

8. Thinking about partitional clustering algorithms (multiple answer):

(a) DBSCAN defines a core point as the point that is at a distance less than a threshold value from another point.
(b) Self Organizing Maps is a message passing algorithm that determines the examples that represents the clusters in the data.
(c) Spectral Clustering is based on the Laplacian Matrix obtained from the similarity matrix of the examples.
(d) Affinity Propagation Clustering is a message passing algorithm that obtains the examples that represent the clusters and how the rest of the examples are assigned to these representatives.

9. Thinking about partitional clustering algorithms (multiple answer):

(a) DBSCAN is not tolerant to noise, unlike K-means.
(b) DBSCAN defines that one example is Direct Density Reachable from a second one if the first belongs to the ϵ neighborhood of the second and the second one is a core point.
(c) Spectral Clustering can be defined as a graph partitioning problem using as weights the values of the Laplacian Matrix and minimizing the weights of the edges of the partitions.
(d) DBSCAN is a grid based clustering algorithm that partitions the space of instances in multidimensional a grid.

10. Thinking about the K-means algorithm and its variants (multiple answer):

(a) The K-means algorithm is able to find arbitrarily shaped clusters in data.
(b) K-medoids is an alternative to K-means that uses one example of the cluster as prototype (medoid) instead of the centroid as K-means does.
(c) The K-means++ initialization strategy for K-means chooses as initial centroids a random sample of size k of the examples.
(d) Bisecting K-means is a variant of the K-means algorithm that begins with one cluster and iteratively adds a new cluster by splitting one of the existing clusters chosen by a quality criteria until the k clusters are obtained.

11. Thinking about the K-means algorithm and its variants (multiple answer):

(a) K-means always returns the same clusters independently of their initialization.
(b) Gobal K-means is a variant of the K-means algorithm that begins with one cluster and iterativelly adds a new cluster considering all the examples as new centroids and picking the best one.
(c) The K-means algorithm is sensitive to outliers and clusters of different sizes or densities.
(d) The K-means algorithm is a partitional algorithm that finds the optimal centroids minimizing the square distance among the examples.

12. Thinking about hierarchical clustering (multiple answer):

(a) The single link criteria for hierarchical clustering forms a new cluster whenever a clique that links all the examples of two clusters appears.
(b) The complete link criteria for hierarchical clustering forms a new cluster whenever a new link appears between two examples of disjoint clusters.
(c) ✓ The Category Utility is a heuristical measure that is used by the COBWEB algorithm for deciding the operator to apply to include a new example in the current hierarchy.
(d) ✓ Graph based hierarchical clustering algorithms can be defined as agglomerative or divisive algorithms.

13. Thinking about hierarchical clustering (multiple answer):

(a) Hierarchical clustering algorithms are computationally efficient with a computational complexity linear in the number of instances.
(b) In Matrix Algebra based hierarchical clustering algorithms the distances to a new formed cluster can only be recomputed by using the distance between the closest examples of the clusters.
(c) ✓ The result of hierarchical clustering is a tree (usually binary) that links the examples according their relationship and not a partition of the data.
(d) ✓ COBWEB is a hierarchical clustering algorithm that can incrementally build non binary hierarchies of examples.

14. Thinking about clustering validation (multiple answer):

(a) ✓ The Gap statistic is a criteria used for deciding the more adequate number of clusters for a dataset.
(b) ✓ The Rand statistic and the Jaccard coefficient are based on the counts of the coincidences of pairs of examples for two different clusterings.
(c) External clustering validity indices are defined using the clustering scatter matrices.
(d) From the scatter matrices defined by the distances between examples and prototypes several external clustering quality indices can be defined.

15. Thinking about clustering validation (multiple answer):

(a) ✓ The idea of cluster stability for determining the number of clusters is that comparing the clusterings of different samplings of the data should results in more similar results for the correct k.
(b) ✓ External clustering validity indices only need to know how the examples are assigned to clusters to be computed.
(c) Some external clustering validity indices are computed using the distribution of the distances of the examples around the centroid.
(d) The only way of knowing if there are clusters in a dataset is to apply a clustering algorithm and see what happens.

16. Thinking about clustering validation (multiple answer):

(a) Normalized Mutual Information is an internal clustering validity index based on the counts of the coincidences of pairs of examples for two different clusterings.
(b) ✓ Internal clustering validation indices are based on the model of the clusters and measure properties that a good clustering should have.
(c) ✓ Some internal clustering validity indices can be used to estimate the number of clusters for a dataset.
(d) Davis-Bouldin criteria and Silhouette index are external clustering validity indices.

17. Thinking about clustering in KDD (multiple answer):

(a) ✓ The Compression/Summarization strategy for clustering scalability is based on discarding examples and using sufficient statistics to summarize their information.
(b) The one pass strategy for clustering scalability is based on the compression of examples using sufficient statistics.
(c) The BIRCH algorithm iterativelly samples data from a database and each step decides what examples to keep and what examples to compress in subclusters.
(d) ✓ Approximation strategies for clustering scalability use for instance approximate data structures for computing distances and frequencies or or perform a preprocesses of the data that reduce the information needed to make decisions .

18. Thinking about clustering time series and graphs (multiple answer):

(a) ✓ The landmark window model for clustering data streams assumes that there are certain points of interest in time for the data stream and the model and when one of these points arrives, the data until that point is discarded.
(b) Karger's algorithm is an exact 2-way graph partitional algorithms that computes the optimal solution for the graph bipartitioning.

(c) Fourier transform and Wavelet tranform are feature extraction methods for time series based on the frequency domain.
(d) The computational cost of Dynamic Time Warping is linear on the length of the time series to compare.

19. Thinking about clustering time series and graphs (multiple answer):

(a) The damped window model for clustering data streams assumes that a weight is asociated with examples and clusters that allows fading the influence of older data.
(b) The Girvan-Newman algorithm is based on finding a set of communities in the graph, colapsing them in one node each and then repeating the process until the desired number of communities is found.
(c) Some algorithms for clustering data streams use a two phase algorithm, an off-line phase that obtains a large number of small clusters and a on-line phase that uses another clustering algorithm to obtain the model for the current time window.
(d) The 2-way graph partitioning problem can be applied for finding clusters in large graphs and can be solved in polynomial time.

20. Thinking about clustering time series and graphs (multiple answer):

(a) Clustering of data streams assumes that the data is arriving continuously and a model of its behavior can be obtained by incremental clustering.
(b) The Girvan-Newman algorithm is a graph partitioning algorithm that uses the eigendecomposition of the laplacian matrix of the graph for performing the clustering.
(c) The Kerninghan-Lin algorithm is a k-way graph partitioning algorithm that can obtain the optimal partitioning in polynomial time.
(d) Symbolic Aggregate Approximation (SAX) is a time series feature extraction method that transforms a series to a vocabulary of discrete symbols.

End Quiz