Guaranteed questions

1. **What is machine learning (ML)?**

   A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (such as planning how to collect more data!)

2. **What are the main ML types?**

   The main types are:

   (a) Supervised Learning
   (b) Unsupervised Learning
   (c) Semi-Supervised Learning
   (d) Reinforcement Learning

3. **What ML algorithms have you studied after the midterm exam?**

   (a) K-Means Clustering
   (b) Hierarchical Clustering
   (c) Recommender Systems
   (d) Large Scale and Online Learning
   (e) Ensemble Learning
   (f) k-Nearest Neighbors (kNNs)
   (g) Principle Components Analysis (PCA)
   (h) Recurrent Neural Networks
   (i) Reinforcement Learning
   (j) Autoencoders
   (k) Bayesian Networks

4. **Which is more important to you — model accuracy, or model performance, support your answer with an example? This is a partial answer, you need to provide a simple example and your opinion.**

   The model accuracy, or model performance is based on your opinion supported by a simple example (hint: all answers are correct such as either one or both together based on the example you provide).

5. **What are advantages and disadvantages of the Hidden Markov Model?**

   (25 Bayesian Networks — Thu, Apr 16

L15 K-Means Clustering — Tue Mar 3

1. **List, then define the common clustering algorithms.**

   K-Means clustering: partitions data into k distinct clusters based on distance to the centroid of a cluster.

   Hierarchical clustering: builds a multilevel hierarchy of clusters by creating a cluster tree.

   Gaussian mixture models: models clusters as a mixture of multivariate normal density components.

   Self-organizing maps: use neural networks that learn the topology and distribution of the data.

   Hidden Markov models: use observed data to recover the sequence of states.

2. **What are the two main steps of the k-means algorithm?**

   (a) Assign

   (b) Optimize (Cost Function)

3. **Write the pseudocode of the k-means algorithm.**

   Randomly initialize K cluster centroids $\mu_1, \mu_2, \mu_3, \ldots, \mu_K \in \mathbb{R}^n$
   repeat
   {
   for i = 1 to m
   $c^{(i)}$:= index (from 1 to K) of cluster centroid closest to $x^{(i)}$
   for $k = 1$ to K
   μ_k := average (mean) of points assigned to cluster k
   }

4. **How does the k-means algorithm work?**

   The way k-means algorithm works is as follows:

   (a) Specify number of clusters K.

   (b) Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.

   (c) Keep iterating until there is no change to the centroids (i.e., assignment of data points to clusters isn't changing).

      i. Compute the sum of the squared distance between data points and all centroids.

      ii. Assign each data point to the closest cluster (centroid).

      iii. Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

5. **List advantages and disadvantages of k-means.**

   Advantages

   - Easy to implement

- With a large number of variables, K-Means may be computationally faster than (than what??)
- k-Means may produce tighter clusters than hierarchical clustering
- An instance can change cluster (move to another cluster) when the centroids are recomputed.

Disadvantages

- Difficult to predict the number of clusters (K-Value)
- Initial seeds have a strong impact on the final results
- The order of the data has an impact on the final results
- Sensitive to scale: rescaling your datasets (normalization or standardization) will completely change results. While this itself is not bad, not realizing that you have to spend extra time on to scaling your data might be bad.

L16 Hierarchical Clustering — Thu Mar 5

1. **What is cluster analysis?**

   - Cluster: A collection of data objects
     - similar (or related) to one another within the same group
     - dissimilar (or unrelated) to the objects in other groups
   - Cluster analysis (or clustering, data segmentation, . . .)
     - Finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters
   - Unsupervised learning: no predefined classes (i.e., learning by observations vs. learning by examples: supervised)

2. **What are the typical applications of cluster analysis?**

   - As a stand-alone tool to get insight into data distribution.
   - As a preprocessing step for other algorithms.

3. **List, then define the two approaches of hierarchical clustering.**

   - Agglomerative: a bottom-up strategy
     - Initially each data object is in its own (atomic) cluster.
     - Then merge these atomic clusters into larger and larger clusters.
   - Divisive: a top-down strategy
     - Initially, all objects are in one single cluster.
     - Then the cluster is subdivided into smaller and smaller clusters.

4. **List all steps of the hierarchical clustering of agglomerative (bottom-up) approach.**

   Step 1: Make each data point a single-point cluster $\rightarrow$ That forms N clusters

   Step 2: Take the two closest data points and make them one cluster $\rightarrow$ That forms $N-1$ clusters

   Step 3: Take **the two closest clusters** and make them one cluster $\rightarrow$ That forms $N-2$ clusters

   Step 4: Repeat Step 3 until there is only one cluster

   Step 5: Finish

5. **Define the dendrograms, then illustrate how do dendrograms work with a diagram.**

   A dendrogram is a diagram that shows the hierarchical relationship between objects.

   - A binary tree that shows how clusters are merged/split hierarchically
   - Each node on the tree is a cluster; each leaf node is a singleton cluster

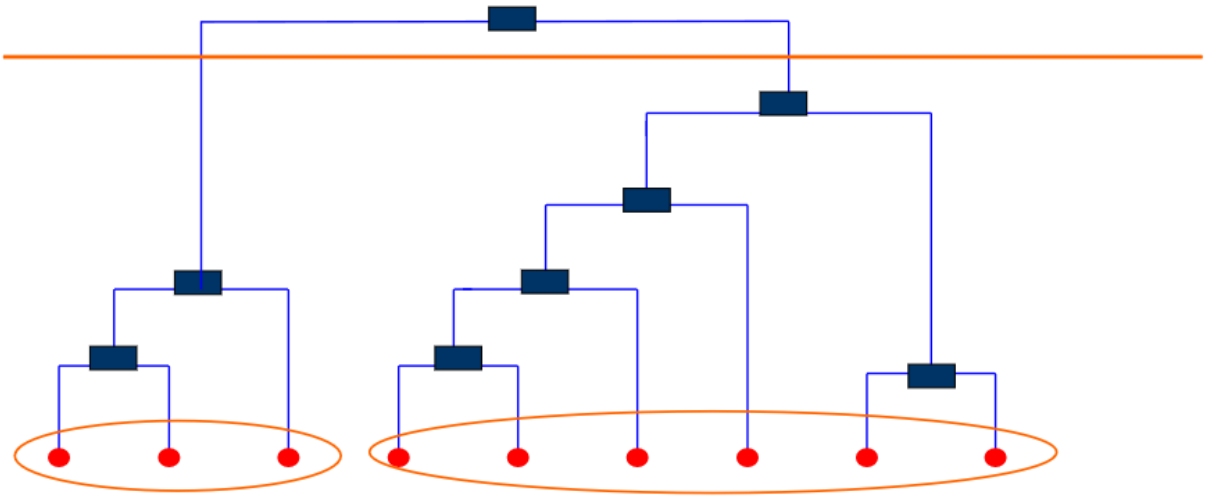

Figure 1: Dendogram

   A clustering of the data objects is obtained by cutting the dendrogram at the desired level, then each connected component forms a cluster

6. **List, then define all possible methods of merging the clusters that depend on the distance measures.**

   Single-link The distance between two clusters is represented by the distance of the **closest pair of data objects** belonging to different clusters.

   Complete-link The distance between two clusters is represented by the distance of the **farthest pair of data objects** belonging to different clusters.

   Average-link The distance between two clusters is represented by the average distance of **all pairs of data objects** belonging to different clusters.

Centroid distance The distance between two clusters is represented by **the means of the clusters**.

7. **What are the advantages and disadvantages of hierarchical clustering?**

   Advantages

   - Hierarchical clustering outputs a hierarchy, i.e., a structure that is more informative than the unstructured set of flat clusters returned by k-means. Therefore, it is easier to decide on the number of clusters by looking at the dendrograms
   - Easy to implement

   Disadvantages

   - It is not possible to undo the previous step: once the instances have been assigned to a cluster, they can no longer be moved around.
   - Time complexity: not suitable for large datasets
   - Initial seeds have a strong impact on the final results
   - The order of the data has an impact on the final results
   - Very sensitive to outliers

L17 Recommender Systems — Tue Mar 10

1. **Define the recommendation systems, why using Recommender Systems?**
2. **What types of recommendation systems, list them, then draw diagrams show the working mechanism of each?**
3. **List advantages and disadvantages of both collaborative filtering and content-based recommenders.**
4. **How to fill rates of users who have not rated any movies?**

L18 Large Scale and Online Learning — Thu Mar 12

1. **Supervised Learning, Semi-Supervised, and Unsupervised Learning for what kinds of applications can be used? What is the different between them in terms of input and output samples?**
2. **What are the differences between Gradient Descent types: Batch, Stochastic, and Mini batch? Which one is the faster to converge?**
3. **What are the hardware-based solutions can be used to machine learning for big data?**
4. **What are the platforms for online machine learning algorithms?**

L19 Ensemble Learning — Thu Mar 19

1. **Define the ensemble learning, illustrate the key motivation of the ensemble learning, then draw the general idea diagram of the ensemble learning**
2. **List the ensemble methods that minimize variance and bias.**
3. **What are the different methods for changing training data? List them, then illustrate the working mechanism of each method, support your working mechanisms with illustration diagrams.**
4. **Can a set of weak learners create a single strong learner?**
5. **What are the main features of the Random Forest method?**

L20 k-Nearest Neighbors (kNNs) — Tue Mar 24

1. **What are the Idea, algorithm, and types of the Instance-Based Learning?**
2. **List the k-Nearest Neighbors (k-NNs) Main Steps.**
3. **What are the three require things to implement the k-NNs?**
4. **How to classify an unknown instance (sample) using the k-NNs?**
5. **What are the two common distance metrics used for k-NNs?**
6. **List Advantages and Disadvantages of k-NNs.**

L21 Principle Components Analysis (PCA) — Thu Mar 26

1. **Define the principle components analysis (PCA), then list the 3 main fields could be used to and 3 application examples.**
2. **What do we mean by the variance and covariance? List the differences between the variance and covariance.**
3. **Illustrate the main tasks of the PCA Process — step 1.**
4. **How we could derive new datasets through the PCA Process — step 5?**

L22 Recurrent Neural Networks — Tue Apr 7

1. **Define RNNs, the show whether RNNs are Supervised or Unsupervised Learning?**
2. **What is the major difference between RNNs and FNNs? illustrate that.**
3. **List types AND architectures of RNNs, then draw the architecture of traditional RNNs.**
4. **List, then illustrate the three main training approaches of RNNs.**
5. **What are the pros and cons of the typical RNNs architecture?**

L23 Reinforcement Learning — Thu Apr 9

1. **List the four main machine learning types.**
2. **Define the reinforcement learning with a diagram, then compare between the reinforcement learning and supervised learning.**
3. **Draw the generic learning model to learn from data. Then define the main operations of it through indicating each operation (i.e. Sensor Data, Feature Extraction, etc.) and related steps.**
4. **What are the key features and elements of the reinforcement learning?**
5. **List the 3 types of reinforcement learning.**
6. **What makes reinforcement learning different from other machine learning paradigms?**

L24 Autoencoders — Tue Apr 14

1. **What are autoencoders? List the general types of autoencoders based on size of hidden layer?**
2. **What are the main differences between PCA and autoencoders?**
3. **List the key elements AND components of autoencoders? Then illustrate the components.**
4. **List, then explain the 3 main properties AND 4 hyperparameters of autoencoders.**
5. **List the 8 types AND 5 applications of autoencoders.**

L25 Bayesian Networks — Thu Apr 16

1. **What are Bayesian networks (BNs)? List BN components and importance.**
2. **List types of probabilistic relationships, then provide 7 real-world Bayesian network applications.**
3. **Define hidden Markov model (HMM), then list and illustrate components of HMM.**
4. **List, with illustration, the 4 main inference algorithms of Hidden Markov Model.**
5. **What are advantages and disadvantages of Hidden Markov Model?**