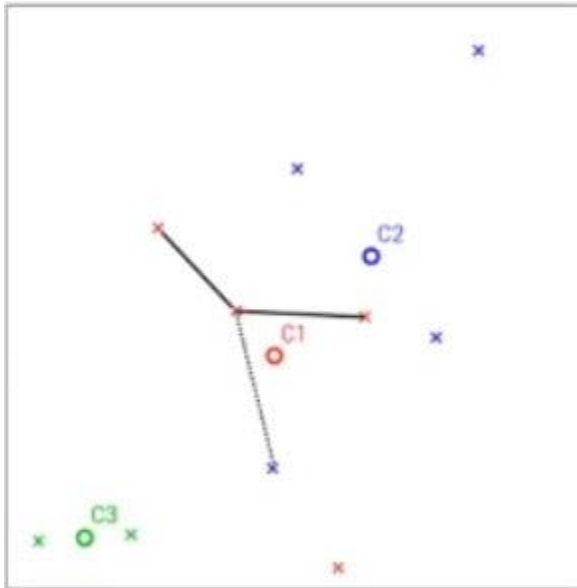


Question 1:



What is this diagram a good example of?

☐ K-nearest neighbor

☐ a decision tree

☐ a linear regression

☒ a K-means cluster

(Correct)

Question 2:

Random forest is modified and improved version of which earlier technique?

☐ aggregated trees

☐ boosted trees

☒ bagged trees (Correct)

☐ stacked trees

Question 3:

Self-organizing maps are specialized neural network for which type of machine learning?

☒ semi-supervised learning (Correct)

☐ supervised learning

☐ reinforcement learning

☐ unsupervised learning

Question 4:

Which statement about K-means clustering is true?

- ☒ In K-means clustering, the initial centroids are sometimes randomly selected. (Correct)
- ☐ K-means clustering is often used in supervised machine learning.
- ☐ The number of clusters are always randomly selected.
- ☐ To be accurate, you want your centroids outside of the cluster.

Question 5:

You created machine learning system that interacts with its environment and responds to errors and rewards. What type of machine learning system is it?

- ☐ supervised learning
- ☐ semi-supervised learning
- ☒ reinforcement learning (Correct)
- ☐ unsupervised learning

Question 6:

Your data science team must build a binary classifier, and the number one criterion is the fastest possible scoring at deployment. It may even be deployed in real time. Which technique will produce a model that will likely be fastest for the deployment team use to new cases?

☐ random forest

☒ logistic regression (Correct)

☐ KNN

☐ deep neural network

Question 7:

Your data science team wants to use the K-nearest neighbor classification algorithm. Someone on your team wants to use a K of 25. What are the challenges of this approach?

☐ Higher K values will produce noisy data.

☐ Higher K values lower the bias but increase the variance.

☐ Higher K values need a larger training set.

☒ Higher K values lower the variance but increase the bias. (Correct)

Question 8:

Your machine learning system is attempting to describe a hidden structure from unlabeled data. How would you describe this machine learning method?

☐ supervised learning

☒ unsupervised learning (Correct)

☐ reinforcement learning

☐ semi-supervised learning

Question 9:

You work for a large credit card processing company that wants to create targeted promotions for its customers. The data science team created a machine learning system that groups together customers who made similar purchases, and divides those customers based on customer loyalty. How would you describe this machine learning approach?

☒ It uses unsupervised learning to cluster together transactions and unsupervised learning to classify the customers. (Correct)

☐ It uses only unsupervised machine learning.

☐ It uses supervised learning to create clusters and unsupervised learning for classification.

☐ It uses reinforcement learning to classify the customers.

Question 10:

You are using K-nearest neighbor and you have a K of 1. What are you likely to see when you train the model?

- ☒ high variance and low bias (Correct)
- ☐ low bias and low variance
- ☐ low variance and high bias
- ☐ high bias and high variance

Question 11:

Are data model bias and variance a challenge with unsupervised learning?

- ☐ No, data model bias and variance are only a challenge with reinforcement learning.
- ☒ Yes, data model bias is a challenge when the machine creates clusters. (Correct)
- ☐ Yes, data model variance trains the unsupervised machine learning algorithm.
- ☐ No, data model bias and variance involve supervised learning.

Question 12:

Which choice is best for binary classification?

☐ K-means

☒ Logistic regression

(Correct)

☐ Linear regression

☐ Principal Component Analysis (PCA)

Explanation

Logistic regression is far better than linear regression at binary classification since it biases the result toward one extreme or the other. K-means clustering can be used for classification but is not as accurate in most scenarios.

Question 13:

With traditional programming, the programmer typically inputs commands. With machine learning, the programmer inputs

☐ supervised learning

☒ data

(Correct)

☐ unsupervised learning

☐ algorithms

Explanation

<https://towardsdatascience.com/machine-learning-for-beginners-d247a9420dab>

Question 14:

Why is it important for machine learning algorithms to have access to high-quality data?

- ☐ It will take too long for programmers to scrub poor data.
- ☐ If the data is high quality, the algorithms will be easier to develop.
- ☐ Low-quality data requires much more processing power than high-quality data.
- ☒ If the data is low quality, you will get inaccurate results. (Correct)

Question 15:

In K-nearest neighbor, the closer you are to neighbor, the more likely you are to

- ☒ share common characteristics (Correct)
- ☐ be part of the root node
- ☐ have a Euclidean connection
- ☐ be part of the same cluster

Question 16:

In the HBO show Silicon Valley, one of the characters creates a mobile application called Not Hot Dog. It works by having the user take a photograph of food with their mobile device. Then the app says whether the food is a hot dog. To create the app, the software developer uploaded hundreds of thousands of pictures of hot dogs. How would you describe this type of machine learning?

☐ Reinforcement machine learning

☐ unsupervised machine learning

☒ supervised machine learning

(Correct)

☐ semi-supervised machine learning

Question 17:

You work for a large pharmaceutical company whose data science team wants to use unsupervised learning machine algorithms to help discover new drugs. What is an advantage to this approach?

- ☐ You will be able to prioritize different classes of drugs, such as antibiotics.
- ☐ You can create a training set of drugs you would like to discover.
- ☒ The algorithms will cluster together drugs that have similar traits. (Correct)
- ☐ Human experts can create classes of drugs to help guide discovery.

Question 18:

In 2015, Google created a machine learning system that could beat a human in the game of Go. This extremely complex game is thought to have more gameplay possibilities than there are atoms of the universe. The first version of the system won by observing hundreds of thousands of hours of human gameplay; the second version learned how to play by getting rewards while playing against itself. How would you describe this transition to different machine learning approaches?

- ☒ The system went from supervised learning to reinforcement learning. (Correct)
- ☐ The system evolved from supervised learning to unsupervised learning.
- ☐ The system evolved from unsupervised learning to supervised learning.
- ☐ The system evolved from reinforcement learning to unsupervised learning.

Question 19:

The security company you work for is thinking about adding machine learning algorithms to their computer network threat detection appliance. What is one advantage of using machine learning?

☒ It could better protect against undiscovered threats. (Correct)

☐ It would very likely lower the hardware requirements.

☐ It would substantially shorten your development time.

☐ It would increase the speed of the appliance.

Question 20:

You work for a hospital that is tracking the community spread of a virus. The hospital created a smartwatch application that uploads body temperature data from hundreds of thousands of participants. What is the best technique to analyze the data?

☐ Use reinforcement learning to reward the system when a new person participates.

☐ Use unsupervised machine learning to cluster together people based on patterns the machine discovers.

☐ Use Supervised machine learning to sort people by demographic data.

☒ Use Supervised machine learning to classify people by body temperature. (Correct)

Question 21:

Many of the advances in machine learning have come from improved ____.

☐ statistics

☐ structured data

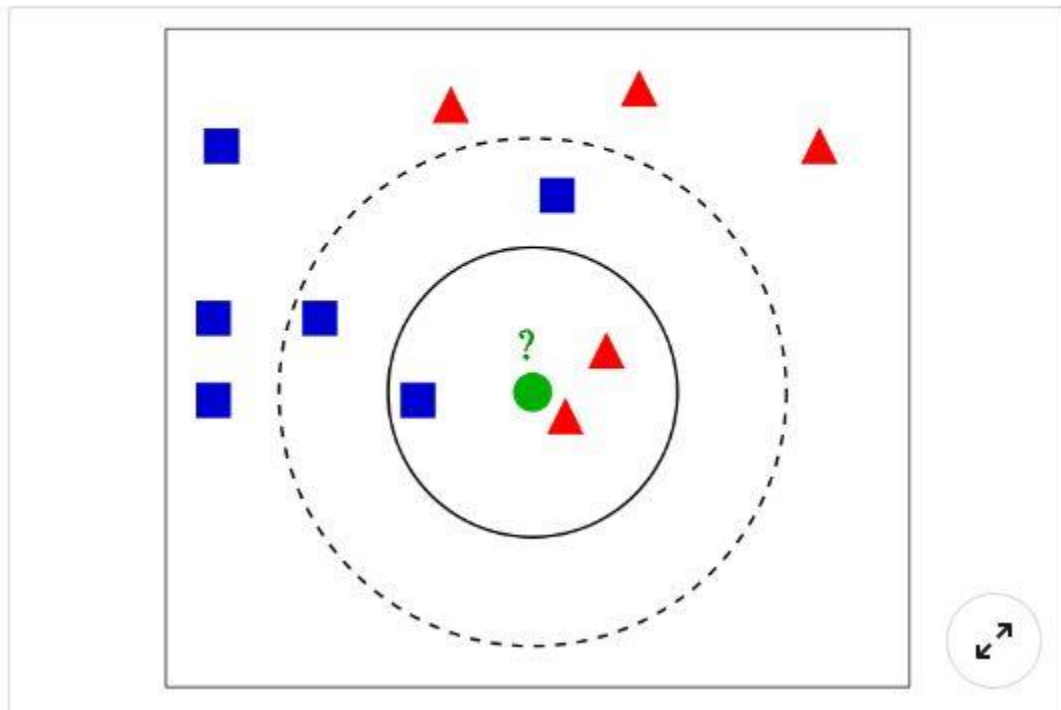
☐ availability

☒ algorithms

(Correct)

Question 22:

What is this diagram a good example of?



☐ unsupervised learning

☐ complex cluster

☐ multiclass classification

☒ k-nearest neighbour

(Correct)

Question 23:

Naive Bayes looks at each _ predictor and creates a probability that belongs in each class.

☐ conditional

☐ multiclass

☒ independent

(Correct)

☐ binary

Explanation

<https://towardsdatascience.com/all-about-naive-bayes-8e13cef044cf>

Question 24:

Someone of your data science team recommends that you use decision trees, naive Bayes and K-nearest neighbor, all at the same time, on the same training data, and then average the results. What is this an example of?

☐ regression analysis

☐ unsupervised learning

☐ high -variance modeling

☒ ensemble modeling

(Correct)

Question 25:

Your data science team wants to use machine learning to better filter out spam messages. The team has gathered a database of 100,000 messages that have been identified as spam or not spam. If you are using supervised machine learning, what would you call this data set?

☐ machine learning algorithm

☒ training set

(Correct)

☐ big data test set

☐ data cluster