Question 1:

What is ensemble modeling?

○ when you create an ensemble of your training and test data set

○ when you create an ensemble of different servers to run the algorithms

○ when you find the one best algorithm for your ensemble

◉ when you use several ensembles of machine learning algorithms   (Correct)

Question 2:

What is the best definition for bias in your data model?

○ Bias is when your predicted values are scattered.

◉ Bias is the gap between your predicted value and the outcome.   (Correct)

○ Bias is when your data is wrong for differnet reasons.

○ Bias is when your values are always off by the same percentage.

Question 3:

Which project might be best suited for supervised machine learning?

- ○ data scrubbing
- ◉ predicting a risk score **(Correct)**
- ○ tax filing software
- ○ spreadsheet consolidation

Question 4:

When is a decision tree most commonly used?

- ○ with big data products
- ◉ for supervised machine learning binary classification challenges **(Correct)**
- ○ to find thd best data cluster
- ○ to determine "Q" in Q-learning reinforcement learning

Question 5:

An organisation that owns dozens of shopping malls wants to create a machine learning product that will use facial recognition to identify customers. What is the main challenge of developing such a model?

○ most machine learning models are not designed for video

◉ it might be unethical for the business to identify people without their consent   **(Correct)**

○ it will be difficult to decide between supervised and unsupervised learning

○ the image in the video would not be high quality enough to identify individuals

Question 6:

Identify the model which is trained with data in only a single batch.

○ Offline Learning

○ Batch Learning

◉ Both Answers   **(Correct)**

○ None of the answers

Question 7:

Identify whether true or false: In PCA the number of input dimensions is equal to principal components.

⦿ TRUE	(Correct)

◯ FALSE

Question 8:

Full form of PAC is _____

◯ Probably Approx Cost

⦿ Probably Approximate Correct	(Correct)

◯ Probability Approx Communication

◯ Probably Approximate Computation

Question 9:

What does K stand for in K mean algorithm?

O   Number of Clusters

O   Number of Data

O   Number of attributes

◉   Number of iterations   (Correct)

Question 10:

Which choice is the best example of labeled data?

◉   a spreadsheet   (Correct)

O   20,000 recorded voicemail messages

O   100,000 images of automobiles

O   hundreds of gigabytes of audio files

Question 11:

In statistics, what is defined as the probability of a hypothesis test of finding an effect - if there is an effect to be found?

- ○ confidence

- ○ alpha

- ⦿ power **(Correct)**

- ○ significance

Question 12:

You want to create a machine learning algorithm to identify food recipes on the web. To do this, you create an algorithm that looks at different conditional probabilities. So if the post includes the word flour, it has a slightly stronger probability of being a recipe. If it contains both flour and sugar, it even more likely a recipe. What type of algorithm are you using?

- ⦿ naive Bayes classifier **(Correct)**

- ○ K-nearest neighbor

- ○ multiclass classification

- ○ decision tree

Question 13:

What is lazy learning?

- ○ when the machine learning algorithms do most of the programming

- ○ when you don't do any data scrubbing

- ● when the learning happens continuously   **(Correct)**

- ○ when you run your computation in one big instance at the beginning

Question 14:

What is Q-learning reinforcement learning?

- ○ supervised machine learning with rewards

- ○ a type of unsupervised learning that relies heavily on a well-established model

- ○ a type of reinforcement learning where accuracy degrades over time

- ● a type of reinforcement learning that focuses on rewards   **(Correct)**

Explanation

Q-learning is a model-free reinforcement learning algorithm.Q-learning is a values-based learning algorithm. Value based algorithms updates the value function based on an equation(particularly Bellman equation

Question 15:

The data in your model has low bias and low variance. How would you expect the data points to be grouped together on the diagram?

⦿ They would be grouped tightly together in the predicted outcome. (Correct)

◯ They would be grouped tightly together but far from the predicted.

◯ They would be scattered around the predict outcome.

◯ They would be scattered far away from the predicted outcome.

Explanation

https://medium.com/30-days-of-machine-learning/day-3-k-nearest-neighbors-and-bias-variance-tradeoff-75f84d515bdb

Question 16:

Your machine learning system is using labeled examples to try to predict future data, compare that data to the predicted result, and then the model. What is the best description of this machine learning method?

◯ unsupervised learning

◯ semi-supervised learning

⦿ supervised learning   (Correct)

◯ semi-reinforcement learning

Explanation

Question 17:

In the 1983 movie WarGames, the computer learns how to master the game of chess by playing against itself. What machine learning method was the computer using?

○ binary learning	

○ supervised learning	

○ unsupervised learning	

● reinforcement learning	(Correct)

Explanation

Question 18:

You are working with your machine learning algorithm on something called class predictor probability. What algorithm are you most likely using?

- ○ multiclass binary classification

- ● naive Bayes   (Correct)

- ○ unsupervised classification

- ○ decision tree analysis

Explanation

you could use a naïve Bayes algorithm, to differentiate three classes of dog breeds — terrier, hound, and sport dogs. Each class has three predictors — hair length, height, and weight. The algorithm does something called class predictor probability.

Question 19:

What is one of the most effective way to correct for underfitting your model to the data?

- ○ Create training clusters

- ○ Remove predictors

- ○ Use reinforcement learning

- ● Add more predictors   (Correct)

Question 20:

Your data science team is often criticized for creating reports that are boring or too obvious. What could you do to help improve the team?

- ◉ Suggest that the team is probably underfitting the model to the data. **(Correct)**

- ○ Suggest that unsupervised learning will lead to more interesting results.

- ○ Make sure that they are picking the correct machine learning algorithms.

- ○ Encourage the team to ask more interesting questions.

Question 21:

What is the difference between unstructured and structured data?

- ○ Unstructured data is always text.

- ○ Unstructured data is much easier to store.

- ◉ Structured data has clearly defined data types. **(Correct)**

- ○ Sturctured data is much more popular.

Question 22:

You work for a startup that is trying to develop a software tool that will scan the internet for pictures of people using specific tools. The chief executive is very interested in using machine learning algorithms. What would you recommend as the best place to start?

○ Using an unsupervised machine learning algorithm to cluster together all the photographs.

○ Crate a data lake with an unsupervised machine learning algorithm.

○ Use a combination of unsupervised and supervised machine learning to create machine-defined data clusters.

◉ Use supervised machine learning to classify photographs based on a predetermined training set. **(Correct)**

Question 23:

In supervised machine learning, data scientist often have the challenge of balancing between underfitting or overfitting their data model. They often have to adjust the training set to make better predictions. What is this balance called?

○ the under/over challenge

○ balance between clustering classification

◉ bias-variance trade-off **(Correct)**

○ the multiclass training set challenge

Question 24:

What is conditional probability?

- ◉ the probability that doing one thing has an impact on another thing — (Correct)

- ○ the probability that certain conditions are met

- ○ the probability that, based on certain conditions, something will always be incorrect

- ○ the probability of something being the correct answer

Question 25:

K-means clustering is what type of machine learning algorithm?

- ○ reinforcement

- ○ supervised

- ◉ unsupervised — (Correct)

- ○ classification