1.What does one mean by the term “machine learning”?

Ans: Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.IBM has a rich history with machine learning. One of its own, Arthur Samuel, is credited for coining the term, “machine learning” with his research (PDF, 481 KB) (link resides outside IBM) around the game of checkers. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat almost seems trivial, but it’s considered a major milestone within the field of artificial intelligence. Over the next couple of decades, the technological developments around storage and processing power will enable some innovative products that we know and love today, such as Netflix’s recommendation engine or self-driving cars.Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to ans Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. IBM has a rich history with machine learning. One of its own, Arthur Samuel, is credited for coining the term, “machine learning”. Robert Nealey, the self-proclaimed checkers master, played the game on an IBM 7094 computer in 1962, and he lost to the computer. Compared to what can be done today, this feat almost seems trivial, but it’s considered a major milestone within the field of artificial intelligence. Over the next couple of decades, the technological developments around storage and processing power will enable some innovative products that we know and love today, such as Netflix’s recommendation engine or self-driving cars.Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

2. Can you think of 4 distinct issues where it shines?

Ans: Machine learning algorithms have had good results on problems such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehiclesFour main challenges in Machine Learning include overfitting the data (using a model too complicated), underfitting the data (using a simple model), lacking in data and nonrepresentative data. If your model performs great on the training data but generalizes poorly to new instances

3. What is a labelled training set, and how does it work?

Ans: It’s no secret that machine learning success is derived from the availability of labeled data in the form of a training set and test set that are used by the learning algorithm. The separation of the data into a training portion and a test portion is the way the algorithm learns. You split up the data containing known response variable values into two pieces. The training set is used to train the algorithm, and then you use the trained model on the test set to predict the response variable values that are already known. The final step is to compare the predicted responses against the actual (observed) responses to see how close they are. The difference is the test error metric. Depending on the test error, you can go back to refine the model and repeat the process until you’re satisfied with the accuracy. Labels are the values of the response variables (what’s being predicted) that are used by the algorithm along with the feature variables (predictors). One consistent problem faced by data scientists is how to obtain labels for a given data set for use with machine learning. In this article we’ll see a variety of techniques used down in the trenches. To start things off, here is a nice summary infographic that outlines a number of data set labeling methods

4. What are the two most important tasks that are supervised?

Ans:. The two most common supervised tasks are regression and classification. Common unsupervised tasks include clustering, visualization, dimensionality reduction, and association rule learning.

Supervised learning problems can be further grouped into regression and classification problems.

Classification: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.

Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Some common types of problems built on top of classification and regression include recommendation and time series prediction respectively.

Some popular examples of supervised machine learning algorithms are:

Linear regression for regression problems.

Random forest for classification and regression problems.

Support vector machines for classification problems.

5. Can you think of four examples of unsupervised tasks?

Ans:. Unsupervised Learning

On the other hand, unsupervised learning is suitable for problems that require the algorithms to identify and extract similarities between the inputs so that similar inputs can be categorised together. In contrast to supervised learning, unsupervised learning methods are suitable when the output variables (i.e the labels) are not provided.

The two fundamental types of unsupervised learning methods are clustering and density estimation. The former (which is probably the most commonly used) involves problems where we need to group the data into specific categories (known as clusters) while the latter involves summarizing the distribution of the data.

Unsupervised Learning — Source: Author

Some examples of unsupervised learning algorithms include K-Means Clustering, Principal Component Analysis and Hierarchical Clustering.

6. State the machine learning model that would be best to make a robot walk through

Various unfamiliar terrains?

Ans:. The best Machine Learning algorithm to allow a robot to walk in unknown terrain is Reinforced Learning, where the robot can learn from response of the terrain to optimize itself.Basically, machine learning is the process of training an AI model to make it intelligent enough to perform specific tasks or some varied actions. And to feed the ML algorithms, a set of data is used at a large scale to make sure AI models like robotics can perform precisely.Using deep reinforcement learning, a type of machine learning that borrows from concepts used in psychology, the scientists could avoid hard-programming every walking-related command as well as avoid simulation tests.

7.. Which algorithm will you use to divide your customers into different groups?

Ans:. We will use the k-means clustering algorithm to derive the optimum number of clusters and understand the underlying customer segments based on the data provided.K-means clustering is a popular unsupervised machine learning algorithm method. In layman terms, it finds all of the different “clusters” and groups them together while keeping them as small as possible. That means that you end up with the most possible customer segments to interpretIn a business context: Clustering algorithm is a technique that assists customer segmentation which is a process of classifying similar customers into the same segment. Clustering algorithm helps to better understand customers, in terms of both static demographics and dynamic behaviors. One of the key challenges that marketing teams must solve is allocating their resources in a way that minimizes “cost per acquisition” (CPA) and increases return on investment. This is possible through segmentation, the process of dividing customers into different groups based on their behavior or characteristics.

Customer segmentation can help reduce waste in marketing campaigns. If you know which customers are similar to each other, you’ll be better positioned to target your campaigns at the right people.

Customer segmentation can also help in other marketing tasks such as product recommendations, pricing, and up-selling strategies.

Customer segmentation was previously a challenging and time-consuming task, that demanded hours of manually poring over different tables and querying the data in hopes of finding ways to group customers together. But in recent years, it has become much easier thanks to machine learning, artificial intelligence algorithms that find statistical regularities in data. Machine learning models can process customer data and discover recurring patterns across various features. In many cases, machine learning algorithms can help marketing analysts find customer segments that would be very difficult to spot through intuition and manual examination of data.

8. Will you consider the problem of spam detectio learning problem? A supervised or unsupervised.?

Ans:. Spam detection is a supervised machine learning problem. This means you must provide your machine learning model with a set of examples of spam and ham messages and let it find the relevant patterns that separate the two different categories. Most email providers have their own vast data sets of labeled emails.The machine learning model used by Google have now advanced to the point that it can detect and filter out spam and phishing emails with about 99.9 percent accuracy. The implication of this is that one out of a thousand messages succeed in evading their email spam filter.Spam detection is messy. The line between spam and non-spam messages is fuzzy, and the criteria change over time. From various efforts to automate spam detection, machine learning has so far proven to be the most effective and the favored approach by email providers. Although we still see spammy emails, a quick look at the junk folder will show how much spam gets weeded out of our inboxes every day thanks to machine learning algorithms.

machine learning determine which emails are spam and which are not Here's an overview of how machine learning–based spam detection works.

9. What is the concept of an online learning system?

Ans: What is Online LearningWhat is Online Learning

1.A term to describe an emerging approach to learn at students' own premise through advanced information-communication technologies (such as Blackboard, Moodle, YouTube, Virtual Reality) either asynchronously or synchronously. Learn more in: What Can College Teachers Learn From Students' Experiential Narratives in Hybrid Courses?: A Text Mining Method of Longitudinal Data2.The type of instruction that is mediated via the internet. Instruction may be synchronous or asynchronous and various technologies can be use to mediate the process. (Dabbagh and Bannan-Ritland, 2005). Learn more in: Constructivist Instructional Design: A Blueprint for Online Course Design3.Learning through the internet via devices that have internet access. Learn more in: Online Course Design Tips for Boosting Learner Autonomy With Synchronous and Asynchronous Tools4.Anytime/anywhere access to education made available through the internet. Learn more in: UDL in Action: Implementing Strategies in a Large Online Course5.Education made available through the internet. Learn more in: Establishing Considerations for Universal Design for Learning and Accessibility in Online Courses6.A teaching mode that does not include in-person sessions and may include synchronous, asynchronous, or hybrid delivery of instruction with some asynchronous and some synchronous online sessions. Learn more in: Pivoting to Deeper Experiences in Education7.The term online learning implies "that the learner is at a distance from the tutor or instructor, that the learner uses some form of technology to access learning materials" (Anderson, 2008). Learn more in: Teaching and Learning Online: An Examination of Effective Techniques, Practices, and Processe

1.A term to describe an emerging approach to learn at students' own premise through advanced information-communication technologies (such as Blackboard, Moodle, YouTube, Virtual Reality) either asynchronously or synchronously. Learn more in: What Can College Teachers Learn From Students' Experiential Narratives in Hybrid Courses?: A Text Mining Method of Longitudinal Data

2.The type of instruction that is mediated via the internet. Instruction may be synchronous or asynchronous and various technologies can be use to mediate the process. (Dabbagh and Bannan-Ritland, 2005). Learn more in: Constructivist Instructional Design: A Blueprint for Online Course Design

3.Learning through the internet via devices that have internet access. Learn more in: Online Course Design Tips for Boosting Learner Autonomy With Synchronous and Asynchronous Tools

4.Anytime/anywhere access to education made available through the internet. Learn more in: UDL in Action: Implementing Strategies in a Large Online Course

5.Education made available through the internet. Learn more in: Establishing Considerations for Universal Design for Learning and Accessibility in Online Courses

6.A teaching mode that does not include in-person sessions and may include synchronous, asynchronous, or hybrid delivery of instruction with some asynchronous and some synchronous online sessions. Learn more in: Pivoting to Deeper Experiences in Education

7.The term online learning implies "that the learner is at a distance from the tutor or instructor, that the learner uses some form of technology to access learning materials" (Anderson, 2008). Learn more in: Teaching and Learning Online: An Examination of Effective Techniques, Practices, and Processes

10. What is out-of-core learning, and how does it differ from core learning?

Ans:. Out-of-core

The term out-of-core typically refers to processing data that is too large to fit into a computer's main memory.

Typically, when a dataset fits neatly into a computer's main memory, randomly accessing sections of data has a (relatively) small performance penalty.

When data must be stored in a medium like a large spinning hard drive or an external computer network, it becomes very expensive to randomly seek to an arbitrary section of data or to process the same data multiple times.

In such a case, an out-of-core algorithm would try to access all relevant data in one sequence.

However, modern computers have a deep memory hierarchy, and replacing random access with sequential access can increase performance even on datasets that fit within memory.Out-of-core learning refers to a set of algorithms working with data that cannot fit into the memory of a single computer, but that can easily fit into some data storage such as a local hard disk or web repository. Your available RAM, the core memory on your single machine, may indeed range from a few gigabytes (sometimes 2 GB, more commonly 4 GB, but we assume that you have 2 GB at maximum) up to 256 GB on large server machines. Large servers are like the ones you can get on cloud computing services such as Amazon Elastic Compute Cloud (EC2), whereas your storage capabilities can easily exceed terabytes of capacity using just an external drive (most likely about 1 TB but it can reach up to 4 TB).

As machine learning is based on globally reducing a cost function, many algorithms initially have been thought to work using all the available data and having access to it at each iteration of the optimization process. This is particularly true for all algorithms based on statistical…

11. What kind of learning algorithm makes predictions using a similarity.

Measure?

Ans:. Learning algorithm that relies on a similarity measure to make predictions is instance-based algorithm.1 — Linear Regression

Linear regression is perhaps one of the most well-known and well-understood algorithms in statistics and machine learning.

Predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. We will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.

The representation of linear regression is an equation that describes a line that best fits the relationship between the input variables (x) and the output variables (y), by finding specific weightings for the input variables called coefficients (B).

Linear Regression

Linear Regression

For example: y = B0 + B1 * x

We will predict y given the input x and the goal of the linear regression learning algorithm is to find the values for the coefficients B0 and B1.

Different techniques can be used to learn the linear regression model from data, such as a linear algebra solution for ordinary least squares and gradient descent optimization.

Linear regression has been around for more than 200 years and has been extensively studied. Some good rules of thumb when using this technique are to remove variables that are very similar (correlated) and to remove noise from your data, if possible. It is a fast and simple technique and good first algorithm to try.

12. What’s the difference between model parameter and a hyperparameter in a learing algorithm.?

Ans:. The data during model training and model hyperparameters are set manually and are used in processes to help estimate model parameters.

Model hyperparameters are often referred to as parameters because they are the parts of the machine learning that must be set manually and tuned.

Basically, parameters are the ones that the “model” uses to make predictions etc. For example, the weight coefficients in a linear regression model. Hyperparameters are the ones that help with the learning process. For example, number of clusters in K-Means, shrinkage factor in Ridge Regression. They won’t appear in the final prediction piece, but they have a large influence on how the parameters would look like after the learning step.Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning. The prefix ‘hyper_’ suggests that they are 'top-level’ parameters that control the learning process and the model parameters that result from it.What is a parameter in a machine learning learning model? A model parameter is a configuration variable that is internal to the model and whose value can be estimated from the given data. They are required by the model when making predictions. Their values define the skill of the model on your problem

13: What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?

Ans:. The goal for a model-based algorithm is to be able to generalize to new examples. To do this, model based algorithms search for optimal values for the model’s parameters, often called theta . This searching, or “learning”, is what machine learning is all about.

Model based learning algorithm search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function.

Recall and Precision, Sensitivity and Specificity are what you are looking for. These are Accuracy measurement metrics that we use heavily in the industry (where the algorithms should meet customer’s criteria of correctness.