1) What does one mean by the term "machine learning"?

Machine learning is a way for computer programs to improve their performance on a task over time given more data.

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

Machine learning algorithms have had good results on problems such has spam detection in email, cancer diagnosis, fraudulent credit card transactions, and automatically driving vehicles.

3) What is a labeled training set?

A labeled training set is a collection of data where one of the features of the data indicates the class the training example belongs to. A labeled training set is used in supervised learning algorithms.

4) What are the two most common supervised tasks?

The two most common supervised learning tasks are regression and classification. In a regression problem we our prediciton is a scalar value. When we're trying to solve a classification problem, our output is either 1 or 0.

5) Can you name 4 common unsupervised tasks?

Common unsupervised tasks include clustering, visualization, dimensionality reduction and association rule learning.

6) What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?

I would use a reinforcement learning approach.
Reinforcement learning is a system where an "agent" observes the environment, selects and performs actions, then recieves a reward or punishment based on the result of the action. Over time the agent learns by itself what is the most productive strategy.

7) What type of algorithm would you use to segment your customers into multiple groups?

I would use some sort of clustering algorithm that can find the decision boundaries in the groups automatically. This is an unsupervised approach. However, if I already knew the categories of my customers, then I would choose a supervised approach and go with a classification algorithm.

8) Would you frame the problem of spam detection as a supervised learning problem or an unsupervised learning problem?

I would frame it as a supervised learning problem because humans have a general idea about what spam is and what it isn't. We can use this notion to create a labeled dataset for an algorithm to learn from.

9) What is an online learning system?

An online learning system learns from new data on-the-fly. As a result, the system is trained incrementally either by using one example at a time or using a mini-batch approach. This keeps each learning step cheap and memory efficient.

10) What is out-of-core learning?

Out-of-core learning is used when a dataset is too large to fit into a computer's memory. The algorithm loads part of the data, runs a training step, then repeats the process until it has run on all the data.

11) What type of learning algorithm relies on a similarity measure to make predictions?

Instance-based learning algorithms use a measure of similarity to generalise to new cases. In an instance-based learning

system, the algorithm learns the examples by heart, then uses the similarity measure to generalise.

12) What is the difference between a model parameter and a learning algorithm’s hyperparameter?

A hyperparameter is a parameter of the learning algorithm, not the model. For example, in a simple linear regression problem our model is parameterized by theta which is a vector of weights. In order to find the best values for theta we have a cost function which is run repeatedly by the gradient descent algorithm. Gradient descent has a hyperparameter called alpha which is the learning rate of the algorithm.

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

The goal for a model-based algorithm is to be able to generalise 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 system learn by minimising a cost function that measures how bad the system is at making predicitons on

new data, plus a penalty for model complexity if the model is regularized. To make a prediction, a new instance's features are fed into a hypothesis function which uses the minimized theta found by repeatedly running the cost function.