John Fiedler
CSCE 4201

Chapter 1

1. How would you define Machine Learning?
   a. Machine learning can be defined as training a system or computer using data. In other words, you build a system that can learn from data.
2. Can you name four types of problems where it shines?
   a. Problems with no algorithmic solution.
   b. Replace lists of hard-coded rules.
   c. A system that needs to adapt to change or fluctuating environments.
   d. Helping humans learn.
3. What is a labeled training set?
   a. A training set that contains the desired solution for each instance.
4. What are the two most common supervised tasks?
   a. Regression and classification
5. Can you name four common unsupervised tasks?
   a. Clustering
   b. Visualization
   c. Dimensionality reduction
   d. Association rule learning
6. What type of Machine Learning algorithm would you use to allow a robot to walk in various unknown terrains?
   a. The best algorithm in this scenario would likely be reinforcement learning.
7. What type of algorithm would you use to segment your customers into multiple groups?
   a. It depends on how you want to group them. If you know what groups you want then you can use a classification algorithm. If you do not, you can use a clustering algorithm.
8. Would you frame the problem of spam detection as a supervised learning problem or and unsupervised learning problem?
   a. Spam detection would be supervised learning because the spam labels are given at the time of training.
9. What is an online learning system?
   a. An online learning system is one that can learn incrementally and adapt to change.
10. What is out-of-core learning?
    a. Out-of-core learning splits data into multiple batches and uses online learning to train on those batches. It allows the use of large data sets that can't fit onto one computer.
11. What type of learning algorithm relies on a similarity measure to make predictions?
    a. Instance-based learning
12. What is the difference between a model parameter and a learning algorithm's hyperparameter?
    a. Model parameters determine what a model will predict on new instances. A hyperparameter is a parameter of the learning algorithm, not the model.
13. What do model-based learning algorithms search for? What is the most common strategy they use to succeed? How do they make predictions?
    a. Model-based learning algorithms search for an optimal value for model parameters. The most common strategy they use is minimizing a cost function. Finally, to make

  predictions, new instance features are fed into the model prediction function using the parameter values.

14. Can you name four of the main challenges of Machine Learning?
    a. Lack of data
    b. Data quality
    c. Nonrepresentative data
    d. Uninformative features

15. If your model performs great on the training data but generalizes poorly to new instances, what is happening? Can you name three possible solutions?
    a. It is likely overfitting the data in this scenario. Three solutions include, getting more data, simplifying the model, or reducing the noise in the training set.

16. What is a test set, and why would you want to use it?
    a. A test set allows us to estimate the errors that a model will make on new instances. You'd want to use it before you deploy the model.

17. What is the purpose of a validation set?
    a. It allows you to compare multiple models.

18. What is the train-dev set, when do you need it, and how do you use it?
    a. It is used when there may be a mismatch between the training, validation, and testing data sets. It is essentially part of the training set used later. You train a model on the training set and test with the validation set. Then you use the train-dev set.

19. What can go wrong if you tune hyperparameters using the test set?
    a. You could potentially overfit the test set and the generalization error would look more optimistic than it actually is.