**Part 1: Write your answers to each problem**

1. Two common supervised tasks are classification and regression.
2. Two common unsupervised tasks are clustering and association learning.
3. Spam detection will count as unsupervised task because, obviously, it would be dumb to label spam email or call as ‘spam’, but given data from users’ actions towards the certain ‘spam’ as well as common similarities with other spam labeled things, it can detect and label the ‘spam’.
4. Reinforcement learning algorithm would allow a robot to walk in various unknown terrains because it can learn from the action while discovering the various new terrains and apply them while it is walking, which will give it ability to not to stumble.
5. Purpose of validation set is to check if program/function from training set is giving minimal error and ready for test set.
6. Test set is for checking if the function/program from validation set is giving the desired output. Since the function/program is not familiar with the test set, having optimal solution from test set will tell us if task is ready to predict as accurate as possible or not.
7. If your model performs great on the training data but generalizes poorly to new instances, it must have cause of extremely overfitted training set. Three possible solution would be less generalization based on training set or take out irrelevant attributes or remove outliers.
8. I have online bookstore with some users, I assume. With the help of machine learning/data mining, it can create and build model for figuring out which books are trending among which kind of users. Also, it can build what kind of books my certain users are interested in or purchased before therefore, with these models, I can offer books that my users would want and sell as many books as possible to fund my company.
9. First, 20 years ago data mining was not possible as nowadays. Now, everything is going through the internet and it is very easy to obtain someone’s information, and everyone is using so kind social accounts. Second, predicting and finding solution for certain disease or others natural/unnatural requires us to use machine learning.
10. HOLDOUT: one dataset is divided into training set and test set. CROSS VALIDATION: one dataset is divided into fixed datasets and iterate through each set as test set and rest of them used as training set. BOOSTRAP: process of sampling with replacement.

**Part 2: Reflection**

I would give myself 8.5 out of 10. First, my answer to question 3 was not correct as shown on Appendix, but I would still think that it is applicable (-0.5). Second, answer to question 6 was not clear as shown on appendix as well, but it was in right direction (-0.5). Unlike question 3 and 6, I believe that my answer to question 9 is correct, but it feels like missing the point which I cannot figure out what (-0.2). Lastly, I am not sure about my answer to question 10 for bootstrap because even after reading the text many times, I am still having trouble understanding (-0.3). However, I clearly understood holdout and validation techniques and their differences.