**Machine Learning Assignment 1**

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

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

1. **Can you think of 4 distinct types of issues where it shines?**

Machine learning is great for problems whose solution requires a great deal of work or a long list of rules, complex problems that are hard to get a solution of using a traditional method, fluctuating environments and getting insights about complex problems and large data.

1. **What is a labeled training set, and how does it work?**

Labeled training set is a set of training data which has a solution to the problem or task (a.k.a. label).

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.

1. **What are the two most important tasks that are supervised?**

Two most common supervised tasks are classification and regression.

1. **Can you think of four examples of unsupervised tasks?**

Four common unsupervised tasks included clustering, visualization, dimensionality reduction, and association rule learning.

1. **State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?**

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.

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

The best algorithm to segment customers into multiple groups is either supervised learning (if the groups have known labels) or unsupervised learning (if there are no group labels).

1. **Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?**

Spam detection is a supervised learning problem because the labels are known (spam or no spam).

1. **What is the concept of an online learning system?**

Online learning system is a learning system in which the machine learns as data is given in small streams continuously.

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

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.

The Core Machine Learning Framework enables developers to integrate their machine learning models into iOS applications. The underlying technologies powering Core ML are both CPU and GPU. Notably, the machine models run on respective devices allowing local analysis of data. The methods in use include both 'Metal' and 'Accelerate' techniques. In most cases, locally run machine learning models are limited in both complexity and productivity as opposed to cloud-based tools.

1. **What kind of learning algorithm makes predictions using a similarity measure?**

Learning algorithm that relies on a similarity measure to make predictions is instance-based algorithm.

1. **What is the difference between a model parameter and a hyperparameter in a learning algorithm?**

Model parameter determines how a model will predict given a new instance; model usually has more than one parameter (i.e. slope of a linear model). Hyperparameter is a parameter for the learning algorithm, not of a model.

1. **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?**

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. The model makes prediction by using the value of the new instance and the parameters in its function.

1. **Can you name four of the most important Machine Learning challenges?**

Four 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.

1. **What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?**

If the model performs poorly to new instances, then it has overfit on the training data. To solve this, we can do any of the following three: get more data, implement a simpler model, or eliminate outliers or noise from the existing data set.

1. **What exactly is a test set, and why would you need one?**

Test set is a set that you test your model (fit using training data) to see how it performs. Test set is necessary so that you can determine how good (or bad) your model performs.

1. **What is a validation set’s purpose?**

Validation set is a set used to compare between different training models.

1. **What precisely is the train-dev kit, when will you need it, how do you put it to use?**

If you tune hyperparameters using the test sets, then it may not perform well on the out-of-sample data because the model is tuned just for that specific set.

1. **What could go wrong if you use the test set to tune hyperparameters?**

Cross-validation is a tool to compare models without needing a separate validation set. It is preferred over validation set because we can save from breaking of part of the training set to create a validation set, as having more data is valuable regardless.