**Assignment- Machine Learning**

**Instructor: Poonam Saini**

Due Date: 20th Jun, 2020

**Q1:** **Can we model a non-linear relationship with a linear regression? Comment with justification.**

**A1:** The answer is yes, given the relationship should be non-linear monotonic relationship.

**Reason:** It have either ever increasing or ever decreasing slope, which make it possible for linear regression to find a best fit line through the relationship. However if the relationship is non-linear and non-monotonic then it depends on sample. As these relationships have both increasing and decreasing slopes which make it impossible to find a relationship that fits well.

**Example:** Examples of monotonic relationships are logarithmic y=lnx and functions with odd powers such as y=x3. Examples of non-monotonic relationships are functions with even powers y=x2 and trigonometric functions such as y=sinx.

**Q2:** **Determine if the following situation is a classification or regression problem with reason:**

**·         A company wants to launch a new product and wants to know whether it will turn out to be a success or failure. You have information on the last 100 products this company launched, including if it was a success/failure, price, weight, color, and several other variables.**

**A2:** The given problem is of binary classification. Because we want to check whether the product will turn out be a success or failure. The nature of result is discrete value not continuous values.

**Q3: What is feature normalization? Is it essential to run logistic regression? Why or Why not?**

**A3:** Feature normalization is a method used to normalize the range of given features of data.

Here’s the formula for normalization:

**X’ = (X-Xmin)/(Xmax-Xmin)**

Here, Xmax and Xmin are the maximum and the minimum values of the feature respectively.

No, it is not at all essential to run logistic regression. The main reason to use feature normalization is to make the data of same range. This helps in optimisation process. For logistic regression there is an activation function used that is sigmoid function which ensures that the output will only be in between 0 and 1 and based upon it final results are evaluated.

**Q4: Which method is used to best fit the data in logistic regression?**

**A4:** Maximum likelihood method.

**Q5: The following graph plots the error distribution of three different models. Comment on each model in terms of bias-variance: (Fig Reference [Ramesh Johari])**

**A5:** The width of the graph tells us about the variance and the nearness to zero tells us about the bias. Therefore,

1. Bias - Red>Blue>Green
2. Variance - Blue>Green>Red

**Q6: Assume we have a set of data from patients who have visited PGI during the year 2019. A set of features (e.g., temperature, height) have been also extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases). We have decided to use a neural network to solve this problem. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer.**

**A6:** Both are possible.

1. The neural network with a shared hidden layer can capture dependencies between diseases. It can be shown that in some cases when there is a dependency between the output nodes, having a shared node in the hidden layer can improve the accuracy.
2. If there is No dependency between diseases (output Neurons), then we would prefer to have a separate neural network for each disease.

**Q7: *k*-means is deterministic or non-deterministic? Comment.**

**A7:**  k-means is a non-deterministic algo because it might produce different outputs at different times for the same input of data. It all starts with a random k number of clusters and iteratively tries to improve them by applying a search algo to the data. The clusters are determined on the basis of distance of points from the centroids of clusters. Due to this random choice of initial clusters, this may lead to different clusters in the end.

k-means can be converted to a deterministic algorithm by using different initialization methods.

**Q8: What will happen upon choosing a very small value of C (~0) in the SVM?**

**A8:** The value C in SVM refers to the cost of misclassification of data. So if the value of C is very small then misclassification would happen. The classifier can maximize the margin between most of the points, while misclassifying a few points, because the penalty is so low.

**Note: If anyone has enrolled for Coursera (or any MOOC) course on Machine Learning or related topic, mention the details or attach certificate if finished.**

**Programming exercise: (as per ML syllabi)**

**Links mentioned for reference only to implement the following programs-**

• Handwritten digit classification (use SVM)

• Face recognition (Python, OpenCV/OpenFace, dlib)

<https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78>

• Image classification and Object detection

<https://www.analyticsvidhya.com/blog/2019/01/build-image-classification-model-10-minutes/>

• Automated music generation

<https://medium.com/datadriveninvestor/music-generation-using-deep-learning-85010fb982e2>