**Week 1-2 Assignment 1**

*Madeline(Zhiyang) Lin*

Note: Read the resources posted in week 1 folder (the chapter 1 of the eBook and the textbook) before solving the problems.

*Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. by Aurélien Géron | Apr 9, 2017 (Free e-copy online).*

**Question 1:** Give three computer applications for which machine learning approaches seem appropriate and three for which they seem inappropriate. Pick applications that are not already mentioned in our textbook or lecture note, and include a one-sentence justification for each.

**ANS:**

The textbook or lecture note have included the majority of the applications regarding machine learning approaches. Below are the applications that I come up with which may not be mentioned.

Three computer applications for which machine learning approaches seem appropriate:

1. Post/Video/Song Recommendations (i.e. Instagram, Tiktok, Spotify)

- The computer uses algorithms to provide recommendations with the similar features/patterns based on your liked posts/videos/songs

2. Weather Prediction

-A conditional probability event which predicts the weather of the following day based on the weather of previous days

3. Google Translate

-It is a neural machine translation system that uses an artificial neural network to increase fluency and accuracy

Three computer applications for which machine learning approaches seem inappropriate:

1. Art Creation

-Artworks need creation and inspiration from human beings, which computer is unable to replicate

2. Language Sentiment

-Lots of language sentimental analysis are not quite accurate as sentiments are more subjunctive-based

3. E-Diary

-People nowadays use word/e-notebook to write their diaries; those record people’s lives with emotions, which computer is unable to read/understand/replicate

**Question 2:** In this question, only submit a summary of the Article in APA format. You must have short summaries under the headings but do not submit the entire paper. The length of the paper is one to two pages. The name of the article is as follow and is in the week one folder:

**ANS:**

**• How the machine ‘thinks’: Understanding opacity in machine learning algorithm**

Opacity is considered as a problem for socially consequential mechanisms of classification and ranking, such as spam filters, credit card fraud detection, search engines, news trends, market segmentation and advertising, insurance or loan qualification, and credit scoring (Burrell, 2016). This paper distinguishes three forms of opacity, which includes: 1) opacity as intentional corporate or institutional self-protection and concealment and, along with it, the possibility for knowing deception; 2) opacity stemming from the current state of affairs where writing (and reading) code is a specialist skill and; 3) an opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation. The paper will focus on the third.

In terms of the machine learning opacity, the paper gives a few figures of neutral network including: 1) a set of examples of handwritten numbers that a machine learning algorithm can be trained on; 2) a handwritten number in an 8\*8 pixel square; 3) graphical depiction of a neural network; and 4) hidden layer in a neural network. All of these examples illustrate that how to use machine learning in the applications of image recognition and handwriting recognition. However, machine learning is applied to the sorts of problems for which encoding an explicit logic of decision-making functions poorly. In this case, the neural network does not break down handwritten digit recognition into subtasks that are readily intelligible to humans, such as identifying a horizontal bar, a closed oval shape, a diagonal line, etc, which is the limitation.

Next, the paper discusses about the spam filtering, which looks at the automation of a task that calls upon a more conscious form of human reasoning and it relevance to questions of classificatory discrimination. Spam is generally understood to be unwelcome emails, especially those sent in bulk, a designation by network administrators concerned particularly with overtaxing network resources. Spam filtering is an application domain for thinking about machine learning based classification as socially consequential. Messages that are categorized as spam are messages that do not get through to their intended recipients. Support vector machines (“SVM”) are the machine learning model used to implement spam filtering. SVMs are like neural networks and it is essentially a form of linear regression (technically called linear kernel). The “bag of words” approach breaks down texts into atomistic collections of units, words whose ordering is irrelevant. The algorithm surfaces very general terms characteristic of spam emails, often terms that are (in isolation) quite mundane and banal.

Computer scientists term this opacity issue a problem of interpretability. One approach to building more interpretable classifiers is to implement an end-user facing component to provide not only the classification outcome, but also exposing some of the logic of this classification. Additionally, it is often the case that the relationship between a feature and a dimension in the model is not one-to-one. Ways of manipulating dimensionality through principal component analysis (“PCA”) or the kernel trick in SVMs are employed to manage computational constraints or to improve accuracy. For handling this fundamental opacity, there are various proposed approaches. One approach, perhaps surprisingly, is to avoid using machine learning algorithms in certain critical domains of application. There are also ways of simplifying machine learning models such as ‘feature extraction’, an approach that analyses what features actually matter to the classification outcome, removing all other features from the model. Some solutions perhaps wisely abandon answering the ‘why’ question and devise metrics that can, in other ways, evaluate discrimination.

All in all, this paper discusses about three types of opacity and focus on the machine learning opacity. The examples of handwriting recognition and spam filtering helped to illustrate how the workings of machine learning algorithms can escape full understanding and interpretation by humans, even for those with specialized training, even for computer scientists. Finally, it proposed some potential approaches. These approaches will depend upon what a given application space requires.

**Reference**

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning

algorithms. *Big Data & Society.* *3*(1). https://doi.org/10.1177/2053951715622512