Purevmaa Damdinsuren

CPSC 4310/Spring 2020

WR3

**Part 0: Questions from articles**

1. *What are the main challenges and limitations of machine learning in health informatics?*

* The articles mentioned about majority of the studies have been retrospective meaning that it is all based on historical and old data which gives worse results in real life.
* Information on all aspects of a diagnosis or prognosis model can have potential risk of bias and potential usefulness of prediction, so it is like two-sided.
* Sometimes machine learning metrics are represented clinical applicability in a way that is not understandable my clinicians; or it is not good enough to describe what is the model really represent.
* Data are generated within a non-stationary environment with shifting patient population; which cause change in distribution and more.
* Machine learning algorithms use whatever signals are available to achieve the best performance including unknown confounders that may not be reliable which impairs algorithm ability to generalize to other datasets.

1. *Discuss an area or a direction of machine learning community should focus on to overcome the challenges or limitations you answer in 1.*

Information being bias:

* We can generalize the dataset to best represent majority not underrepresented groups.
* Sometimes it probably happens because of the lack of or inadequate dataset, so we can get same datasets to use for model from every individual in same way.
* Other times, it may give issue because it has never experienced with these kind of training data. For example, Russian developers created FaceApp app that has racist effect on black people making them look like white people because their data does not have experience with that.

**Part 1: Questions from textbook**

1. *Discuss about the main difference between Naïve Bayes Classifier and Bayesian Belief Network.*

* Naïve Bayes Classifier performs well only with categorical attributes, and it computes/predicts the result of final target based on the idea that all attributes are conditionally independent to each other.
* Because the assumption of conditional independence is overly restrictive, ‘Bayesian Belief Network all stating conditional independence assumptions that apply to subset of variables,’ so that it is less constraining to learn about its’ model target and it’s tractable.

1. *K-Nearest Neighbor (KNN) classifier is considered as non-parametric technique. What does it mean?*

It does not make assumptions about the model between the class target y and the attributes (x1, x2, … xn). This method considers the similarities between features/attributes of the model and find existing patterns among values instead of estimating parameters from assumption.

1. *Suppose the features in your training set have very different scales. Then the cost function in regression will have the shape of an elongated bowl, so the Gradient Descent algorithms will take a long time to converge. What can we do about it?*

It is better to scale the data before training the model with it. Unscaled target variables on regression problems can cause the learning process to fail.

**Part 2: Write a reflective essay about your learning**

Since I took Artificial Intelligence class last quarter, learning about the topics on Bayesian network was easier than learning it for the first time. However, there were lots of computations to deal with, which we did not learn in AI, so that will be challenging in the later course. In regression part, I think once I follow the given code (within hands-on experience), it will be easy and challenges will be overcame with practices.