

## Part 1:

1. You are asked to create an ML algorithm which can predict whether a patient with phantom limb syndrome will respond to a particular intervention.

- a. Does it make sense to create such an ML algo? Why and/or why not?

Some forms of intervention can be irreversible or really hard to reverse such as invasive surgery. Introducing an ML algorithm that could successfully determine if a particular invasive intervention may work would prevent failure and solve pain sooner for some patients.

- b. If yes, what sort of data would you seek to train the algorithm on? State explicitly what a data object is, what its label (if any) is, and what is the relation that the algorithm will seek to learn. (If not, you don't have to answer this)

The data object in this case would be a patient profile. Standard patient details that would be handed to a doctor would also be included in the data object such as age, gender, time since injury, reason for injury, type of amputation, etc. The label for the data would be if the intervention worked or not.

- c. What kind of ML algorithm would this be: supervised, unsupervised, reinforcement, semi-supervised? (If you answered no to a, you don't have to answer this)

This would be a semi-supervised algorithm. Too much supervision in this case may be restrictive to the algorithm and ask it to perform as a human would which comes with its own set of biases. Instead a semi supervised approach with some labeled data would also for more freedom than a human might apply to the situation.

2. You are asked to create an ML algorithm which can predict whether an article, submitted to a scientific journal, should be accepted for publication or not.

- a. Does it make sense to create such an ML algo? Why and/or why not?

The value of such an algorithm is not very clear. If the purpose is for a researcher to use the algorithm to predict if their article will be accepted, then yes there is value. Any other use of this algorithm, such as by a

reviewer, would only introduce bias into the review process as reviews are supposed to be unbiased. There is a clear-ish rubric that many journals use to determine if a paper should be accepted and data points that can be labeled clearly to do so.

- b. If yes, what sort of data would you seek to train the algorithm on? State explicitly what a data object is, what its label (if any) is, and what is the relation that the algorithm will seek to learn. (If not, you don't have to answer this)**

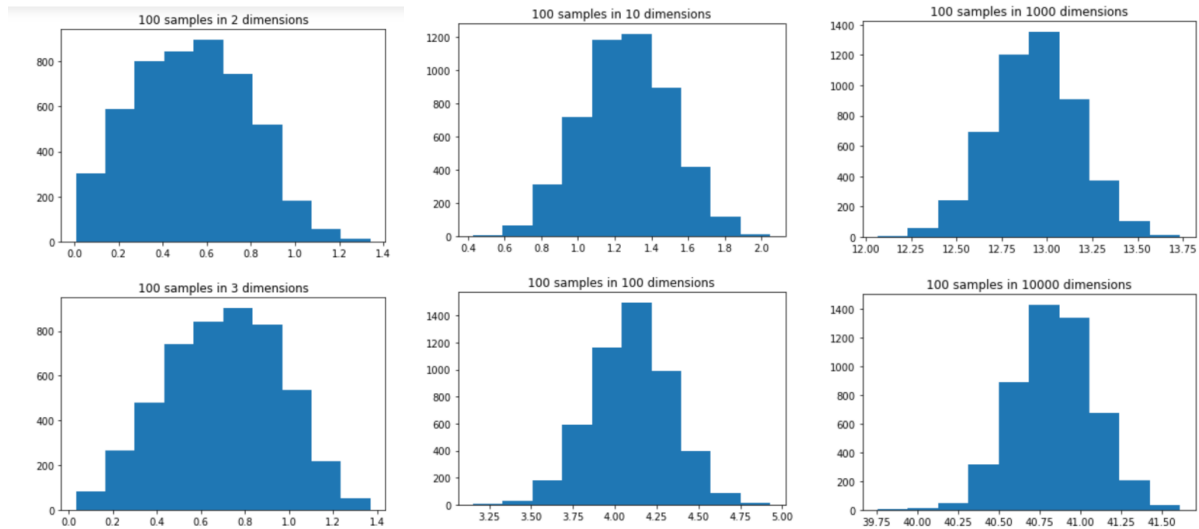
The data would be a set of all papers submitted to a journal within some specified time frame (Ex. past five years) each labeled with if it was accepted or rejected.

- c. What kind of ML algorithm would this be: supervised, unsupervised, reinforcement, semi-supervised? (If you answered no to a, you don't have to answer this)**

There can be arguments for both supervised and unsupervised learning in this case. In the supervised case, we can give a labeled set of data identifying if a paper is accepted or not. There is plenty of available clean data to feed this to an algorithm which makes it more suited for supervised learning.

The unsupervised learning case is interesting because it may identify some features that make a paper more likely to be accepted vs another, however this isn't as useful to the original purpose of identifying a successful paper.

## Part 3:



As  $d$  grows, the distribution of the pairwise distances remains somewhat uniform but starts to become narrower with more of the points being more equidistant from each other. This means as we add more dimensions, the distance between any two points starts to approach the average distance between points.