Big Ideas in AI
ML and Sampling

Thomas Aamand Witting
wittingt@oregonstate.edu

October 31, 2022

1 ML Taxonomy

1.1 Phantom Limb Syndrome

1.1.1 a

It makes sense to attempt to create a machine learning algorithm for this use case as medical data are very comparable in the form of physiological measurements, perceived pain, before and after data etc. This is ideal for labeling algorithms. However, I think it is very important for the result of the algorithm to be reviewed by a doctor specializing in the field, as the algorithm may not take the full context of the patient into account.

1.1.2 b

The training data would consist of anonymized medical and demographic data from patients who has received that particular intervention. The label of the data is whether the patients condition improved after treatment. The algorithm seeks to learn the relation between the medical data and the label, so new patients can receive proper intervention.

1.1.3 c

This is a supervised ML algorithm as it is trained purely on labeled data.

1.2 Admission of Scientific Article

It does not make sense to create a machine learning algorithm for this use case, as it needs to predict whether a proposed articles findings are relevant and evident-based while also predicting whether its finding are valid or not. Training such an algorithm on previous accepted and rejected scientific articles would not make it able to predict the status of future articles, as they likely will contain new information and findings, or even results contradictory to previous articles. Additionally, the possible training data would not be very measurable and comparable between samples, as it almost entirely would consist of text and images which is far from ideal for labeling unseen data.

2 KNN in High Dimensions

In figure 1, the distribution of the pairwise differences from 100 points sampled from the d-dimensional hypercube is seen. As d increases, the distances does as well, requiring an increasing amount of samples to get similar performance as it gets in lower dimensions.

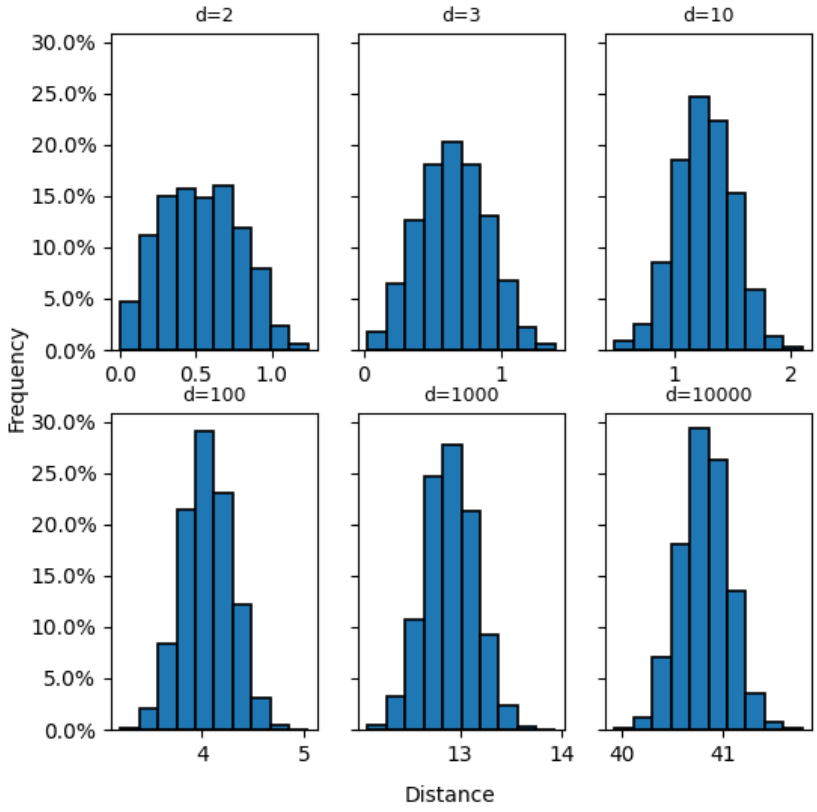

Figure 1: Distance distribution between points (N=100) sampled from a d-dimension hypercube