

Portofolio Assignment 1

Candidate: 25

September 2020

Appendix

Problem 1

(1a)

Supervised learning consists of machine learning algorithms which both has inputs and outputs. The goal of the supervised learning algorithm is using the observed values of x to make an prediction of y , where we, the creator of the algorithm is the supervisor. So generally the algorithm consists of whats called a mapping of x to y . Or we can generalize this as $y = f(x)$, where f is the mapping of x to y . Examples of supervised learning algorithms include classification, regression etc.

Whereas in unsupervised learning the goal of the algorithm is to find connections in the data, such that we can learn more from it. Here we don't have a supervisor, we simply try to better see patterns in the input data. We also aren't interested in any output since the input is used to train the model to rule out differences of the variables.

The PageRank algorithm is a unsupervised learning algorithm.

(1b)

We are given an equation representing the PageRank method,

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|}. \quad (1.1)$$

This sum ranks, with the $r(P_i)$ method, the given page, where p_1, p_2, \dots, p_n represent all the pages we want to compare against, where n is the number of pages. P_j is a page contained within the set of all the other pages that links to P_i , denoted B_{P_i} . Then the same is true for $r(P_j)$ as is for $r(P_i)$. $|P_j|$ represents the number of links from P_j to other pages.

This equation will provide a ranking vector, π , to all the pages p_1, \dots, p_n . That means we need n of these rankings to compute all the ranks for the different pages. If we combine

Problem 2

(2a)

$$y = f(x) \quad (2.1)$$