

Homework -- January 15th, 2019

Jarred Parr

Machine Learning

1. Supervised learning involves the use of training data to tell an algorithm what is right and wrong in the context of a given problem. The training data tells the algorithm what is right and wrong via the use of **class labels**, these class labels tell the algorithm what the correct output is after it has had time to optimize itself on a particular dataset. Unsupervised learning involves the algorithm making an inference about a given dataset. It can find trends and similarities in the data and group data by similar characteristics. This allows scientists to figure out a general shape to the data on a deeper level than just attempting to read it.
 1. Supervised Example: A convolutional neural network learning to recognize hand-written digits
 2. Unsupervised: K-Means clustering on a set of unlabeled data
2. A possible learning task could be in the space of MRI and other medical image reading. The algorithm would read a medical image and detect anomalies from the output that could be overlooked by the cursory glance of a doctor. Of the approaches discussed in section 1.2, classification appears to be the option that is best suited to this task. Most difficulties would come about with the use of confidential patient data, and the low margin for error. An improper diagnosis could be catastrophic for a hospital's reputation and finances. This algorithm would need to also have a significant amount of training data to allow for very optimum tuning to many unique scenarios. False positives would simply not be an option in this context

Statistics And Probability

1. The mean can be calculated as

$$(1 \cdot 20) + (2 \cdot 50) + (3 \cdot 30) / 20 + 30 + 50 = 2.1 \text{ nights on average}$$

2. Since we can only have a valid combination of the four, we have $4!$ options which means there are 24 total situations. With all possible permutations of AGCT we get a total of 8 possible values. As a result, the probability that two purines and two pyrimidines are found together is $8/24$, or 33%.
3. The probability that a patient is female, given that the patient suffers from migraines is the total number of females who have migraines divided by the total number of people that suffer from migraines. Which would be $210/250 = 84\%$
4. The probability that someone has been diagnosed with hepatitis given that they actually have hepatitis is equal to the probability that someone has hepatitis given that they were diagnosed with hepatitis (tested positive), times the probability of having hepatitis, over the probability of testing positive.