Theodore Kim

CS-UY 4563 A – Introduction to Machine Learning

Homework #1

Question 1: *Provide N and d for each data set*

1. Collect a set of data (profit, number of employees, industry and CEO salary) on the top 500 firms to affect CEO salary. **N = 500**, **d = 3**
2. Collect data to determine if a new product is a *success* or *failure*. Collecting data on 20 similar products (success/failure, price, budget, competition price, and ten other variables). **N = 20, d = 13**

Question 2: *Describe a real life applications of the following ML strategies as well as the target and features (explaining inference / prediction)*

1. Classification:
2. A spam email classifier in which the target is whether or not an email is labelled as spam and the features are the sender’s email address, the subject line, the time it was sent and the email message itself. This is an example of a prediction as one would use a dataset of previously identified spam/normal emails with the feature set described previously to build a model to classify future emails
4. Regression:
5. Determining whether one’s max salary is influenced by factors like level of education (i.e. years educated or highest degree), profession, work experience (years), and location (some quantified location, i.e. each state or city gets a number). This is an example of inference as the features are analyzed for whether the outcome fits a linear trend.

Question 3: *A university admissions office wants to predict success of students based on application material*

1. A good target for this supervised learning problem is the alumni’s salaries ten years after graduation.
2. The target variable is continuous
3. One possible predictor for the target variable is the prospective students’ SAT scores
4. A linear model for the relationship would fit, I expect the slope to be positive as higher SAT scores could reasonably mean a higher salary in the future

Question 4: *Data samples (xi, yi)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *xi* | 0 | 1 | 2 | 3 | 4 |
| *yi* | 0 | 2 | 3 | 8 | 17 |

1. Adjusting the training set

Question 5: *Perform 2 steps of gradient descent,*

Question 6: