# Stats Week 9

Rvail Naveed: 17321983

March 24, 2020

## Q1(a)

Code for gradient descent included in zip.

## Q1(b)

### 0.1 $\alpha = 1$

The gradient descent algorithm becomes unstable because the learning rate parameter is too high, causing it to jump between 0.5 and -0.5 constantly. It is not able to come up with an answer in a reasonable timeframe.

## 0.2 $\alpha = 0.1$

The algorithm is able to find the minimum in 32 steps, an adequate amount of time.

- Steps taken: 32
- Minimum occurs at: 0.00

---

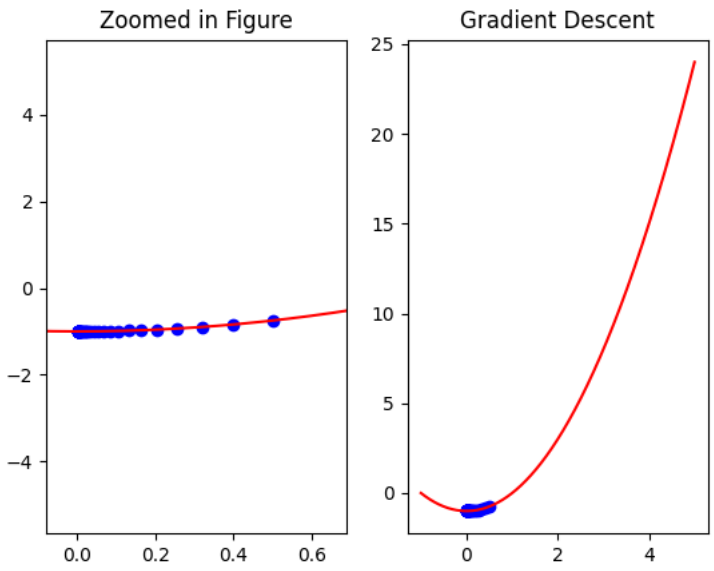



Figure 1: $\alpha = 0.1, x_start = 0.5, precision = 0.0001$

### 0.3 $\alpha = 0.01$

The learning rate is too low, so it takes many more steps to find the minimum. In reality the minimum was probably found before the completion of all steps but the learning rate being too low prevented it from stopping earlier.

- Steps taken: 229
- Minimum occurs at: 0.00

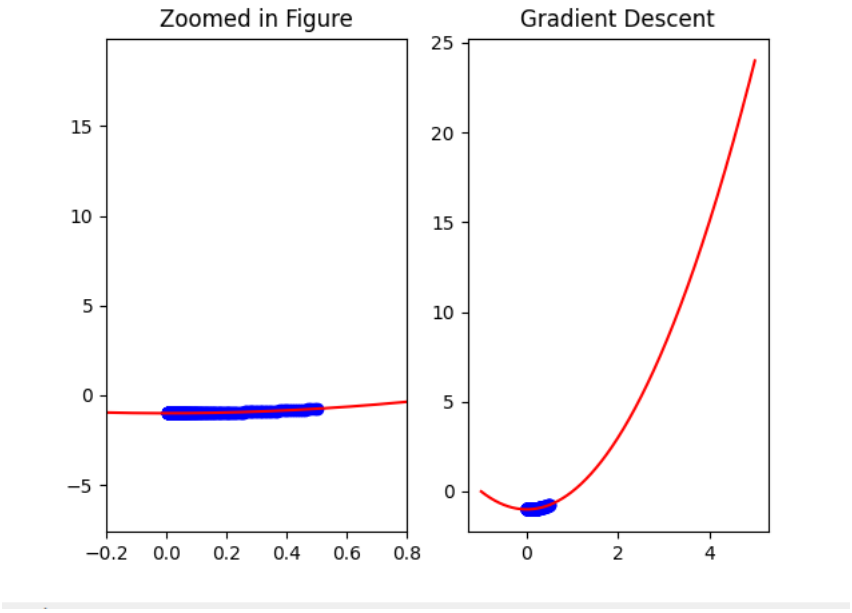
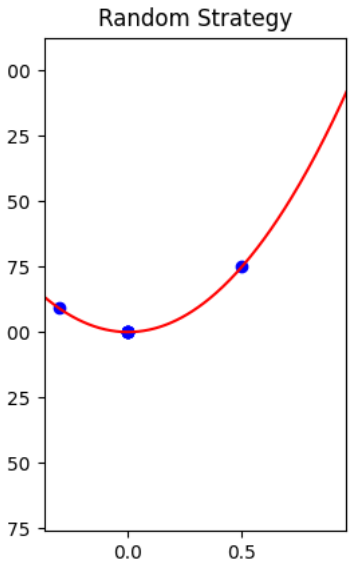



Figure 2: $\alpha = 0.01, x_start = 0.5, precision = 0.0001$

# Q1(c)

Code for random strategy included in zip.
*Only plotted the x values for which $f(x_{k+1}) < f(x)$

- Minimum: 0.00
- Steps: Varies between 5000-50,000




# Q1(d)

Took many more steps to find a suitable x value in the vicinity of the current x. Since picking the x point is random, results vary drastically. This is in stark contrast to gradient descent which is very easy to control via the learning rate.

# Q2(a)

This can be achieved using the Bag of Words approach described in the slides:

- Delete all the little words, truncate word endings.
- Form a list of all the unique words that are left, and number them 1-n.
- Map the text of a review to a vector of length n by setting i of the vector equal to the number of times the ith word in the dictionary appears in the review.

- Then apply logistic regression methods as usual with this new vector.
- Using model: $\frac{P(Y=1|X=x)=1}{1+exp(-z)}$

# Q2(b)

Assumptions:

- There is a linear relationship between the features and target.
- X values in the sample are not all the same (Variability in X values is positive)

# Q2(c)

- Randomly split training data into training set and test set.
- Train model on the training data to select parameter values which minimise cost function.
- Evaluate accuracy of predictions using test data.
- Repeat above steps to obtain a set of prediction errors.
- Use those to estimate distribution and so the confidence interval.