

Jorge Gonzalez

861112270

CS171

4-16-2017

PS 1

Number 1:

Problem Set 1

1.) $P(\text{False-Positive}) = .01$, meaning $P(\text{Positive}) = .99$
 $P(\text{False-Negative}) = .002$, meaning $P(\text{Negative}) = .998$
 $\frac{1}{4000}$ have = .00025

$$P(\text{You have disease} | \text{Positive}) = \frac{P(\text{You have disease})P(\text{Positive} | \text{You have disease})}{P(\text{Positive})}$$
$$= \frac{(.00025)(.99)}{.99}$$

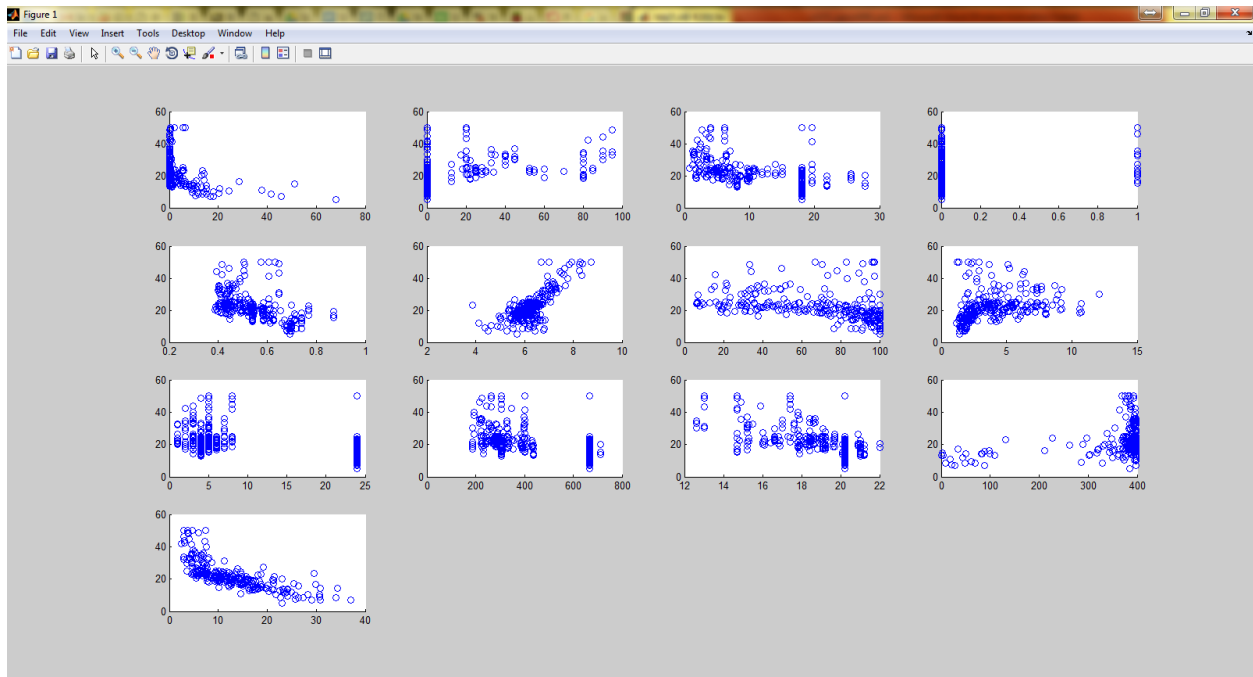
$.99 \cdot .00025 = .0002475$ (people who have disease who test positive)
 $(1 - .00025) \cdot .01 = .0099975$ (people who don't have disease who test positive)

$$\begin{array}{r} .0002475 \\ + .0099975 \\ \hline .0102450 = 1.0245\% \end{array}$$
$$\frac{(.00025)(.99)}{.010245} = .02435$$

= 2.435%, to actually have disease, given a positive test.

Number 2:

check zip file for function `plotdata(fname)`. Image of result below:



Number 3:

```
3. Running plotdata on the housetrain set gave me prediction info on how the
attributes of houses boston affect their housing value. The predictions
are as follows:
→1. As crime rates rise per capita, housing value goes down.
→2. Data is unpredictable in middle values, but as houses are zoned for lots,
→value tends to rise.
→3. As non-retail businesses rise, house values tend to fall.
→4. Housing values tended to have some medium to high prices when near a river
→5. Housing values tended to rise as Nitric oxides neared .5
→6. Housing values rose as number of rooms increased per house
→7. Age of the house did not seem to affect prices of the house as it was
→random, but most houses did stay low priced
→8. prices and sales tended to be more frequent as employer distance decreased.
→9. Prices increase as accessibility to highways became easier
→10. Housing value increased as property tax stayed low
→11. Prices rose high more when ratio stayed between 10 and 20.
→12. As number of blacks in the community increased, housing values rose
→13. As the status of the population decreased, so did housing values
```

Number 4:

Part A: refer to ridgells.m in zip file.

Part B: refer to llerr.m in zip file.

Part C: As far as ridge regression is concerned the plot below tells me that as the function is further plotted, the training data versus the validation data gets more accurate as you add more values to it. As you pick a higher lambda and force the higher degree polynomials to abide by a curve, the more

accurate your results will be.

