CMPT 726 A1 Yu Guo 30126558?

① 1. μ_i for the possibility get i side for $i = 1, 2, 3, 4, 5$

$\mu_6 = 1 - \sum_{i=1}^{N=5} \mu_i$

2. $\mu_i = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$

3. $\mu_2 = 1$ $\mu_1 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = 0$

4. $0 \leq \mu_i \leq 1$ $\sum_{i=1}^{6} \mu_i = 1$ $i = 1, 2, 3, 4, 5, 6$

② $E(w) = \frac{1}{2} \sum_{n=1}^{N} \alpha_n (t_n - w^T \phi(x_n))^2$ want $\nabla E = 0^T$

$$\frac{\partial}{\partial w} E(w) = \sum_{n=1}^{N} \alpha_n (t_n - w^T \phi(x_n)) \phi(x_n)$$

$$0 = \sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T - \sum_{n=1}^{N} \alpha_n w^T \phi(x_n) \phi(x_n)^T$$

$$w^T \sum^{N} \alpha_n \phi(x_n) \phi(x_n)^T = \sum^{N} \alpha_n t_n \phi(x_n)^T$$

$$w^T \alpha X^T X = \alpha^T X^T$$

$$w = (X^T \alpha X)^{-1} X^T \alpha T$$

③ 1, NO, one counter example is if validation data is same as train data, the error will be same, not higher

2 Yes, for unregularized regression, higher degree can make line more close to data, which means lower or same error

3 No, one counter example is for $\lambda = \infty$ in regularized regression. the w will be real small compare to value w should be. So the regulaized error will be higher in this case.

4 $E(w) = \sum_{n=1}^{N} (t_n - w^T \phi(x_n)) + \sum_{i=1}^{i} \lambda_i w^j$

λ and j is hyper parameter input $j = \{1, 2\}$ λ is weight

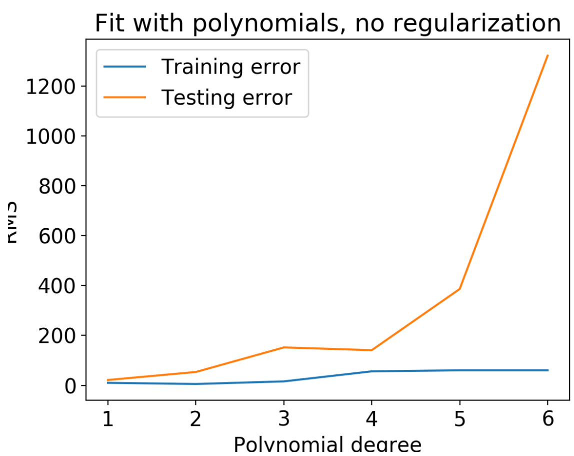
$$\boxed{\nabla E(w) = -\sum_{n=1}^{N} (t_n - w^T \phi(x_n))\, \phi(x_n)^T + j \sum_{i=1}^{i} \lambda_i w^{j-1}}$$

$$\boxed{= -\sum_{n=1}^{N} (t_n - w^T \phi(x_n))\, \phi(x_n)^T + 2 \sum_{1}^{J_1} \lambda_{J_1} w + \sum_{1}^{J_2} \lambda_{J_2}}$$

Problem 5
5.1

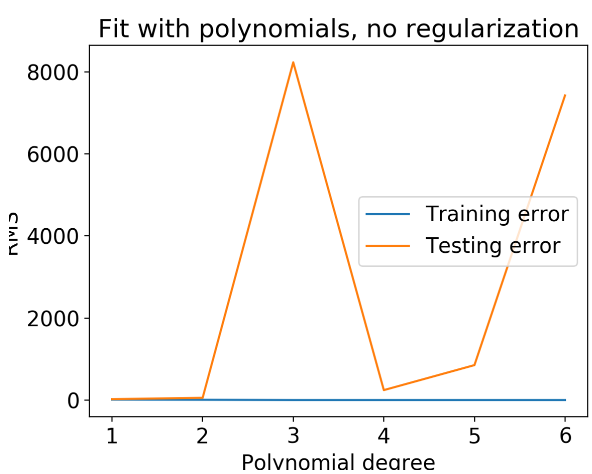
1. Niger - 313.7
2. Sierra Leone - 185.3
3. Replace the missing value with mean in column.

5.2
1.



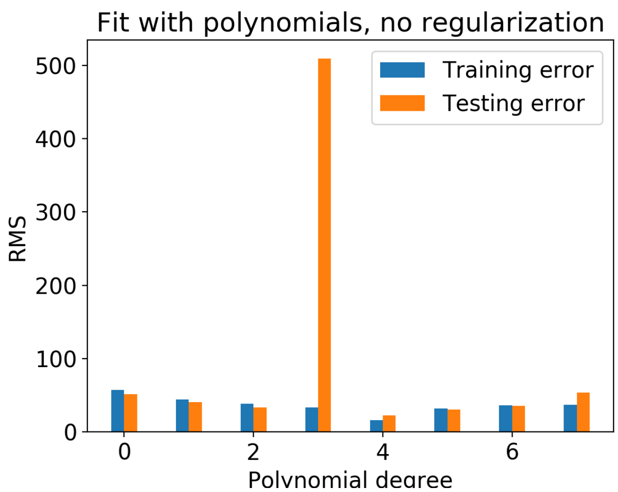
Before normalization

With the increase of polynomial degree, the testing error goes up.

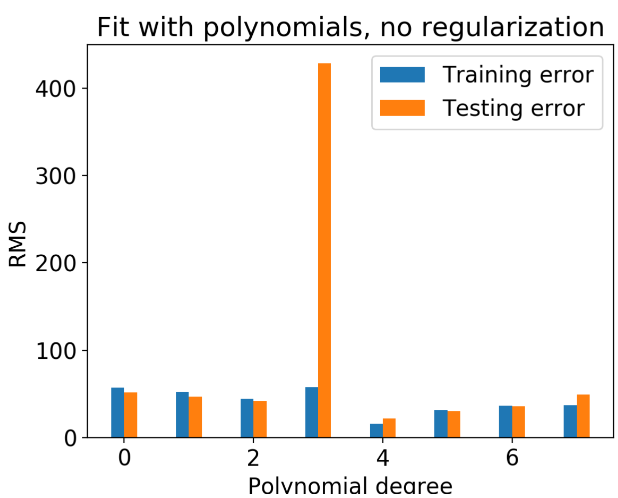


After normalization

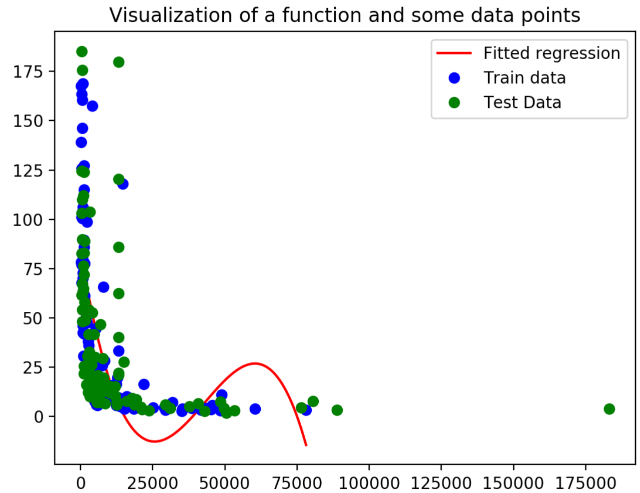
5.2.2



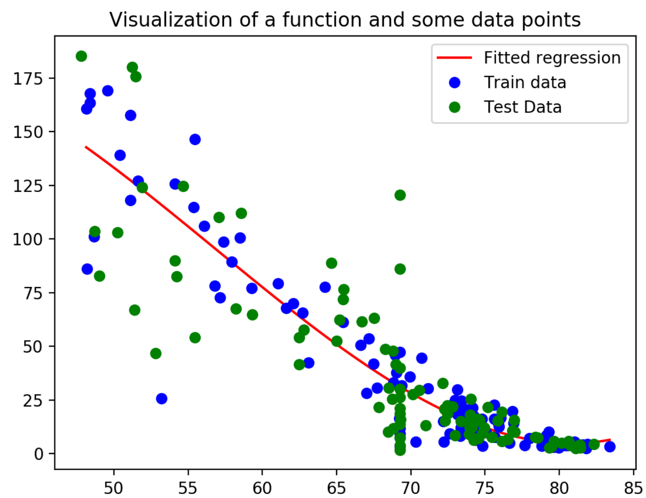
With bias



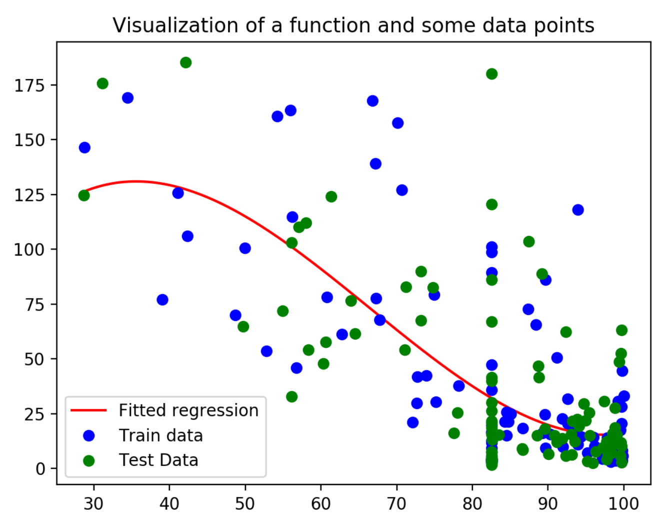
Without bias



GNI

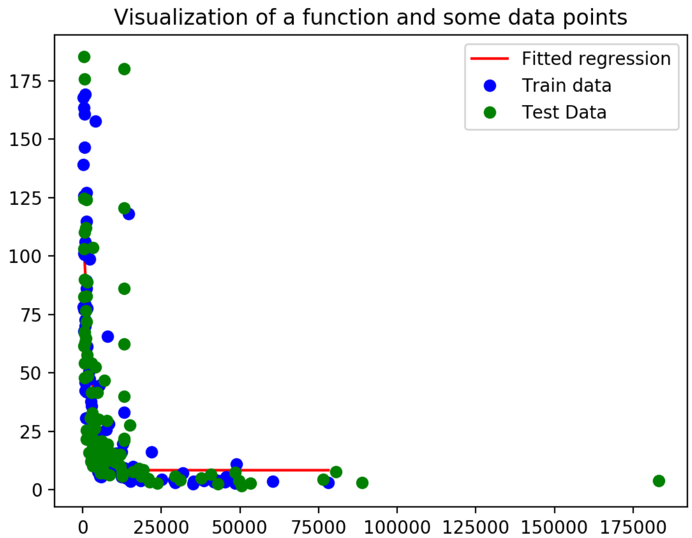


Life expectancy

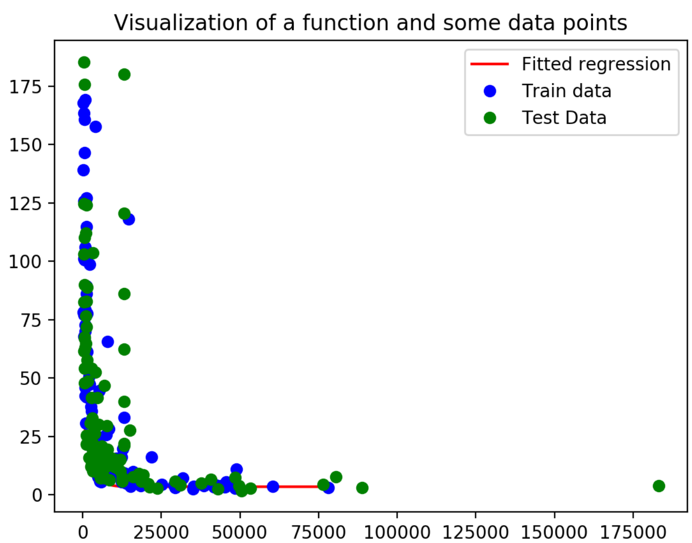


Literacy

5.3

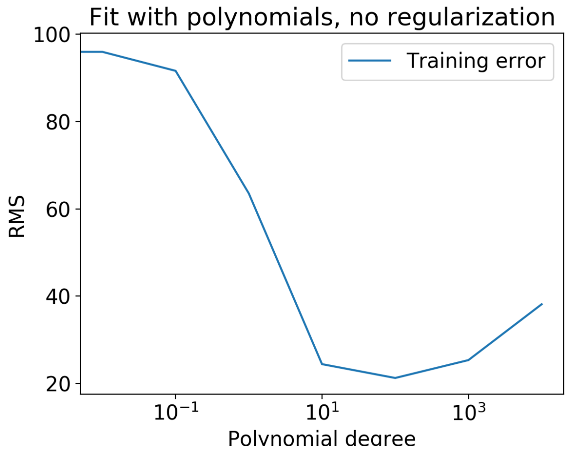


Mu = 100
Train error = 28.555023448223448
Test error = 34.04101841026058



Mu = 10000
Train error = 38.408782047940605
Test error = 39.187056804369604

5.4



Lambda = 100 has the smallest error.

0	134.0872480012051,
0.01	89.93358664559737,
0.1	49.182403187023496,
1	54.03122827800846,
10	38.11164190068872,
10^2	24.056916620748765,
10^3	27.615413925888504,
10^4	44.260654394116614

The 0 and 10^4 don't show up in graph, but in chart