Question 1:

1) $\mu = \{\mu_1, \mu_2, \mu_3\}$ st. $\begin{cases} \mu_1 \text{ is propability of NDP winnig} \\ \mu_2 \quad " \quad \text{Liberal} \\ \mu_3 \quad \text{Green Party} \end{cases}$

$\sum_{i=1}^{3} \mu_i = 1$ and $0 \leq \mu_i \leq 1$

2) $\sum_{i=1}^{3} \mu_i = 1, \; 0 \leq \mu_i \leq 1$ and $\mu_1 = \mu_2 = \mu_3 \Rightarrow \mu_1 = \mu_2 = \mu_3 = 1/3$

3) $\begin{cases} \mu_1 + \mu_2 + \mu_3 = 1 \\ \exists_i \text{ st. } \mu_i = 1 \end{cases} \Rightarrow \{(1,0,0) \text{ or } (0,1,0) \text{ or } (0,0,1)\}$

4) $Dir(\mu|\alpha) = \frac{\Gamma(3\alpha_K)}{\Gamma(\alpha_K)^3} \prod_{K=1}^{3} \mu^{\alpha_K - 1}$ and $\forall_{K=1}^{3} \; 0 < \alpha_K < 1$ and $\alpha_1 = \alpha_2 = \alpha_3$

$\forall_K \; \mu_K = 1/3 \leftarrow \sum_{i=1}^{3} \mu = 1, \; \mu_1 = \mu_2 = \mu_3$

5) $\Gamma(\alpha) = \int_0^\infty u^{\alpha - 1} e^{-u} \, du$

$$P(\mu|D,\alpha) = Dir(\mu|\alpha + m) = \frac{\Gamma(\alpha_0 + N)}{\Gamma(\alpha_1 + m_1) \cdots \Gamma(\alpha_K + m_K)} \prod_{K=1}^{K} \mu_K^{\alpha_K + m_K - 1}$$

Prior (Green Party rigged) $\Rightarrow \alpha = \{1, 1, 0.1\}$ $\Rightarrow P(\mu|D,\alpha) = \frac{\Gamma(2.1+N)}{\Gamma(1+N)\cdot\Gamma(1)\cdot\Gamma(0.1)} \times \left(\frac{1}{3}\right)^N \times \left(\frac{1}{3}\right)^0 \times \left(\frac{1}{3}\right)^{-0.9}$

if there are N polls $\Rightarrow m = \{N, 0, 0\}$

because in all polls, NDP has the largest share of votes $\quad \forall_{K=1}^{3} \mu_K = 1/3$

$$\Rightarrow P(\mu|D,\alpha) = \frac{\Gamma(2.1+N)}{\Gamma(N+1)\cdot\Gamma(1)\cdot\Gamma(0.1)} \times \frac{1}{3}^{(N-0.9)}$$

6. $E(\text{tution}) = \sum_{i=1}^{3} t_i \cdot P(\mu_i|\alpha_i) = t_1 \mu_1^{\alpha_1 - 1} + t_2 \mu_2^{\alpha_2 - 1} + t_3 \mu_3^{\alpha_3 - 1}$

$P(\mu_i|\alpha_i) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_K)} \cdot \mu_i^{\alpha_i - 1}$ $\quad E(x_i) = \frac{\alpha_i}{\alpha_0}$

Question2:

Q2.

$$P(t \mid X, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid w^T \phi(x_n), \beta_n^{-1})$$

$$\ln P(t \mid X, w, \beta) = \sum_{n=1}^{N} \ln \mathcal{N}(t_n \mid w^T \phi(x_n), \beta_n^{-1})$$

$$= \sum_{n=1}^{N} \ln \left(\sqrt{\frac{\beta_n}{2\pi}} \cdot e^{-\frac{\beta_n}{2} (t_n - w^T \phi(x_n))^2} \right)$$

$$= \sum_{n=1}^{N} \frac{\beta_n}{2} - \sum_{n=1}^{N} \pi - \sum_{n=1}^{N} \frac{\beta_n}{2} \cdot (t_n - w^T \phi(x_n))^2$$

$$\Rightarrow \frac{\partial \ln P(t \mid X, w, \beta)}{\partial w} = \underbrace{\frac{\partial}{\partial w}\left(\sum_{n=1}^{N} \frac{\beta_n}{2}\right)}_{0} - \underbrace{\frac{\partial}{\partial w}(N \cdot \pi)}_{0 \text{ constant}} - \frac{\partial}{\partial w}\left(\sum_{n=1}^{N} \frac{\beta_n}{2} \cdot (t_n - w^T \phi(x_n))^2\right)$$

$$\Rightarrow$$

$$= \frac{\partial}{\partial w} \sum_{n=1}^{N} \frac{\beta_n}{2} \cdot (t_n - w^T \phi(x_n))^2$$

$$= \sum_{n=1}^{N} \frac{\beta_n}{\cancel{2}} \times \cancel{2} \, (t_n - w^T \phi(x_n)) \cdot \phi^T(x_n)$$

$$\boxed{= \sum_{n=1}^{N} \beta_n t_n \cdot \phi^T(x_n) - w^T \sum_{n=1}^{N} \beta_n \phi(x_n) \cdot \phi^T(x_n)}$$

Question 3:

1. No. The reason is there might be a training set that is exactly on the trained model by the regression. It means although the assumption is that it is unregularized and the model can over fit, the statement claim that the validation error is "Always" "Higher" than training error, which is wrong. Because both might be zero.

2. Yes. For dataset with 10 or less data points, both model can memorize the data because both are unregolarized. so both error will be zero. Incase of having more than 10 data points polynomial with degree 10 includes degree 9 as well so it can provide lower error on the training data.

3: No. The statements claims Always. Althoug in most scenario it might be true, there might be test set that is exactly on the unregularized regression model, which cause the other one return greater error. So, the answer is No

Question 4:

Q4.

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - w^T\phi(x_n)\}^2 + \overbrace{\sum_{i=1}^{M}\frac{\lambda_i}{2}|w_i|}^{L_1} + \overbrace{\sum_{i=1}^{M}\frac{\lambda_i}{2}|w_2|^2}^{L_2}$$

$$\nabla E(w) = \underbrace{\sum_{n=1}^{N} t_n\,\phi(x_n)^T}_{1} - \underbrace{w^T\sum_{n=1}^{N}\phi(x_n)\phi^T(x_n)}_{2} + \underbrace{\sum_{i=1}^{M}\frac{\lambda_i}{2}\times\frac{|w_i|}{w_i}}_{3} + \underbrace{\sum_{i=1}^{M}\lambda_i\,w_i}_{4}$$

Question 5:

Step 1-z installed numpy, scipy, pandas, ..

```
python -m pip install --user numpy scipy matplotlib ipython jupyter pandas sympy nose
```

5.1. Getting Started:

1- Which country had the highest child mortality rate in 1990? What was the rate?

Country :Niger , Under-5 mortality rate (U5MR) 1990: 313.7

2- Which country had the highest child mortality rate in 2011? What was the rate?

Country: Sierra Leone, Under-5 mortality rate (U5MR) 2011: 185.3

3- Some countries are missing some features (see original .xlsx/.csv spreadsheet). How is this handled in the function assignment1.load unicef data()?

Using the following codes,

1- it finds the arithmetic mean value of each column while ignoring nan values and store it in the mean_value array.

    mean_vals = np.nanmean(values, axis=0)

2- Finds the indexes of the cells with nan value and store it in the ins array.

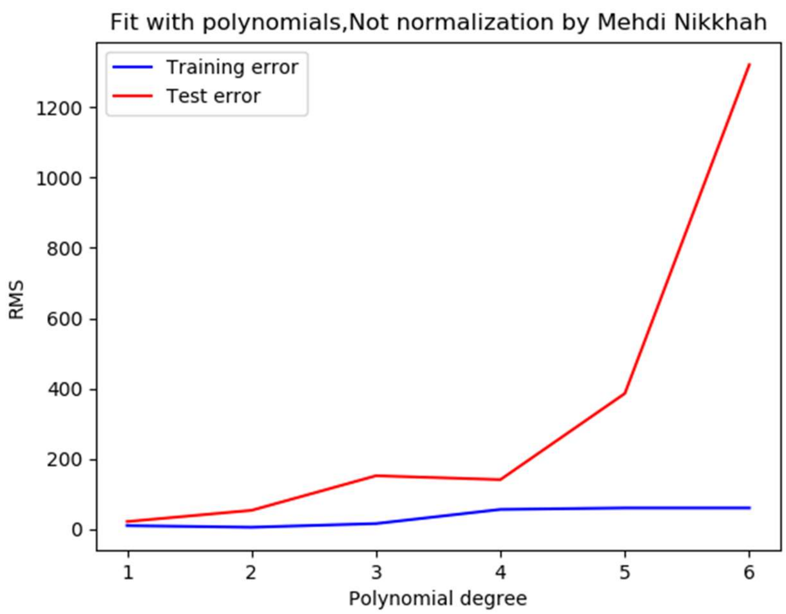
    inds = np.where(np.isnan(values))

3- Replace the nan in the values matrix with the related column's mean_value.

```
values[inds] = np.take(mean_vals, inds[1])
```

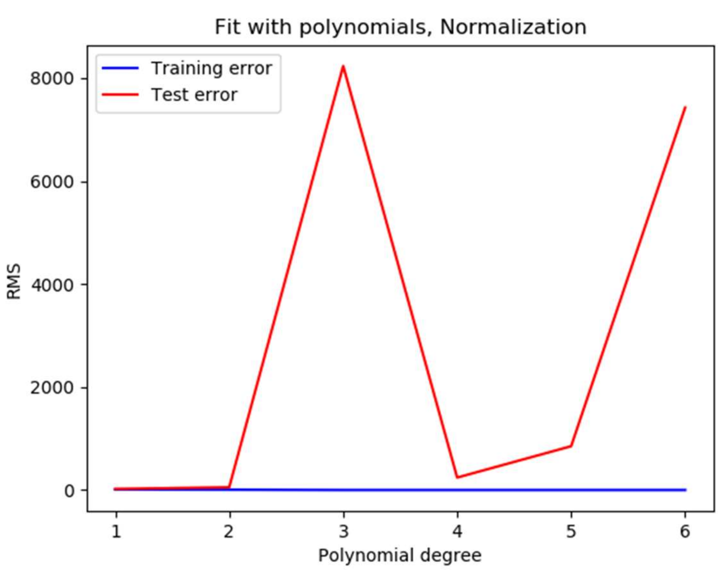
5.2

5.2.1.1



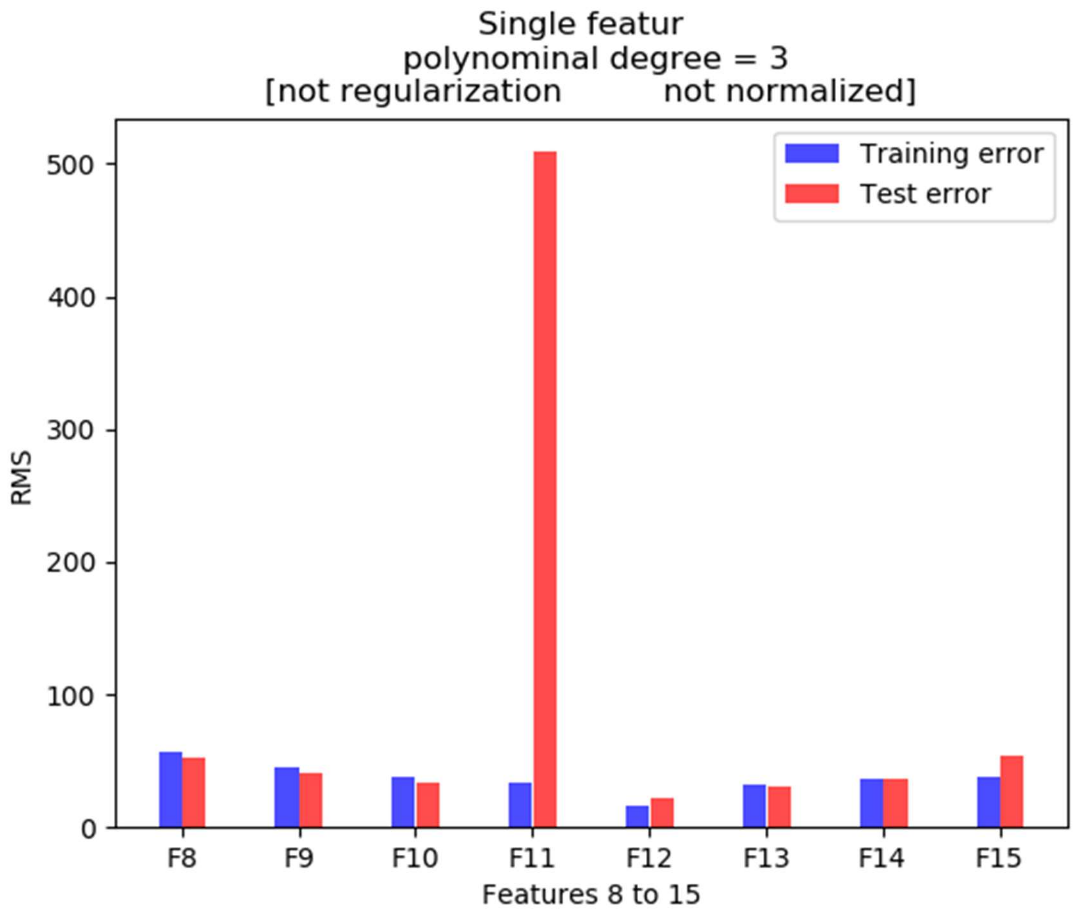
In point 3 we can see the error is higher, because there is a wrong data which cause the square error takes a large value. The reason is that input is not normalized.

5.2.1.2.

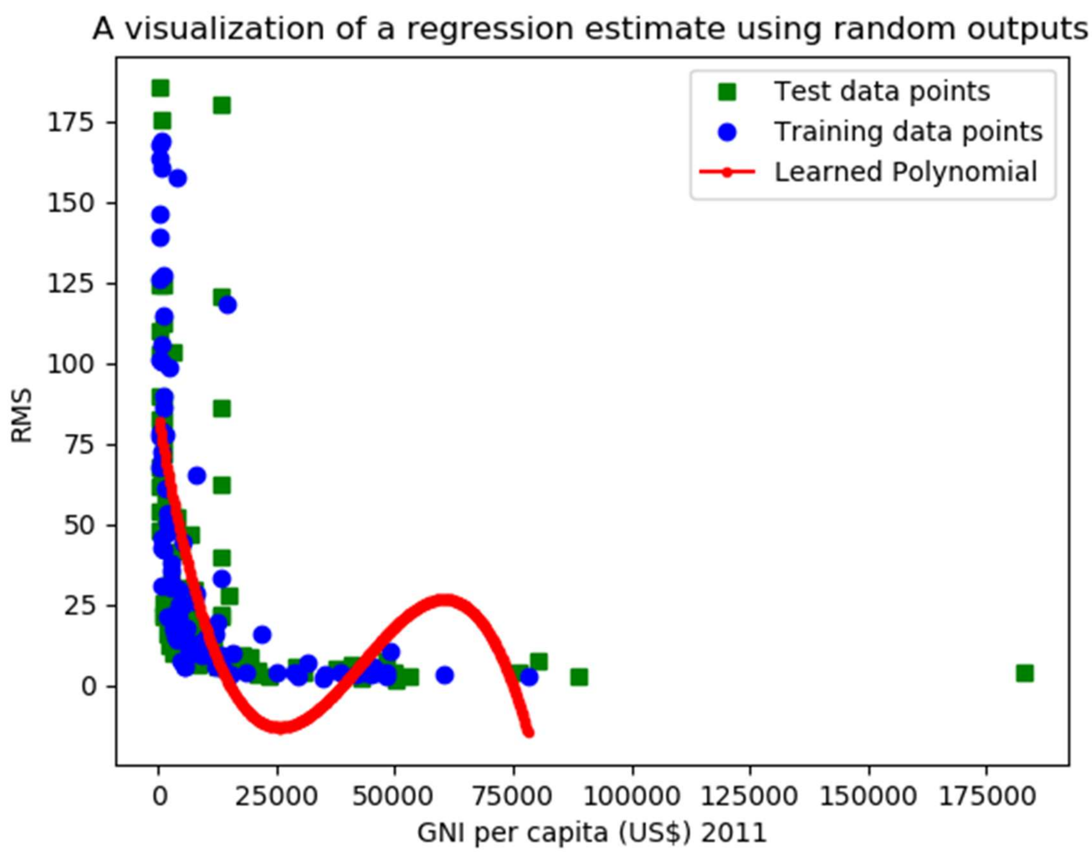


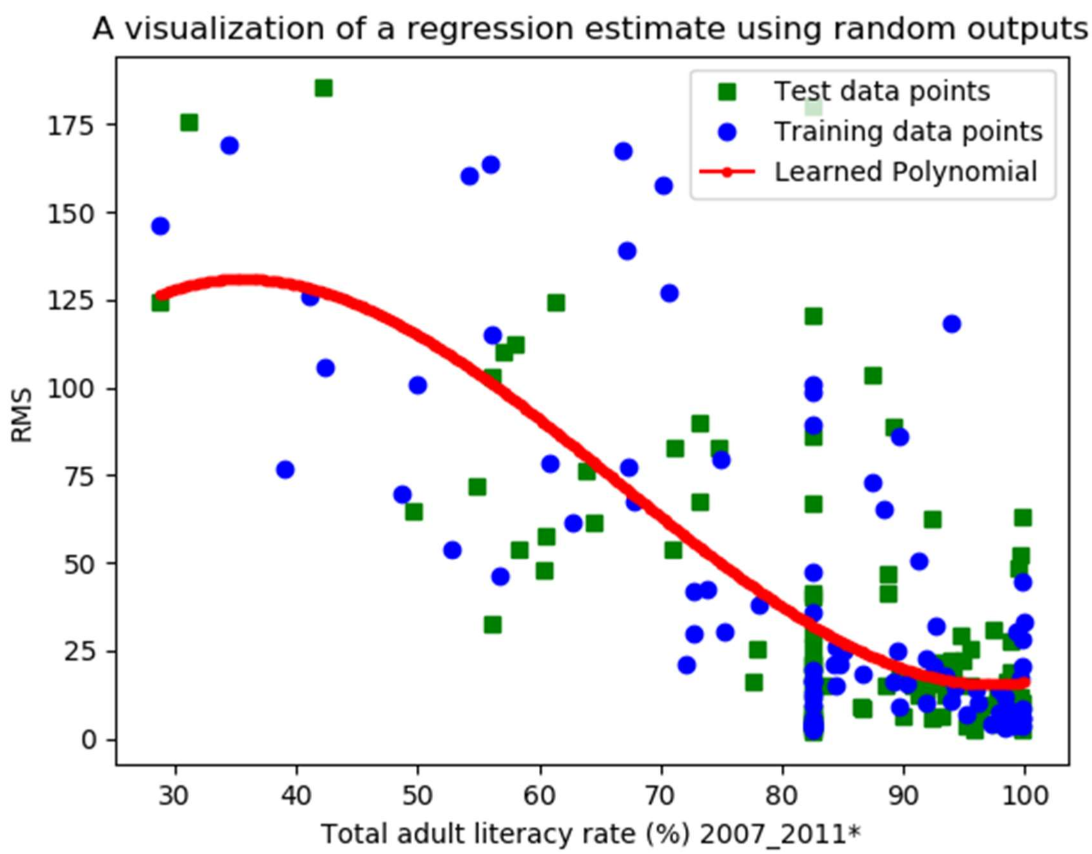
5.2.2.

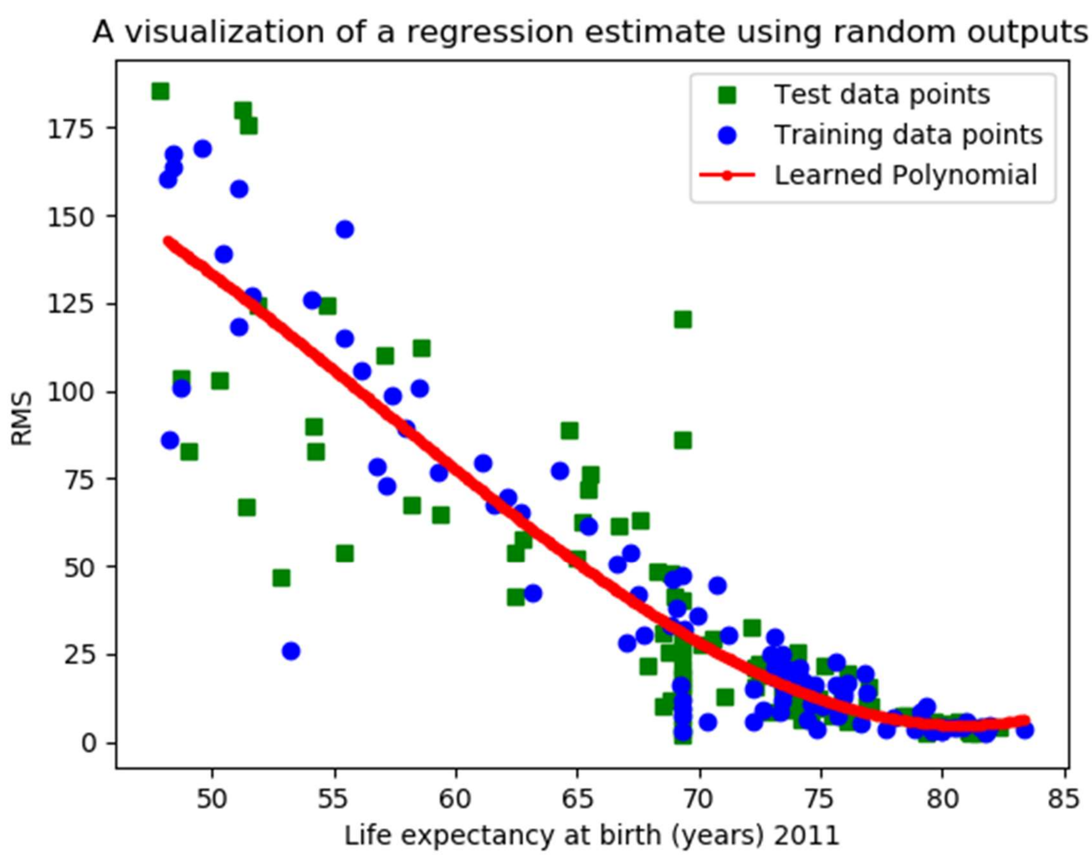
5.2.2.1.



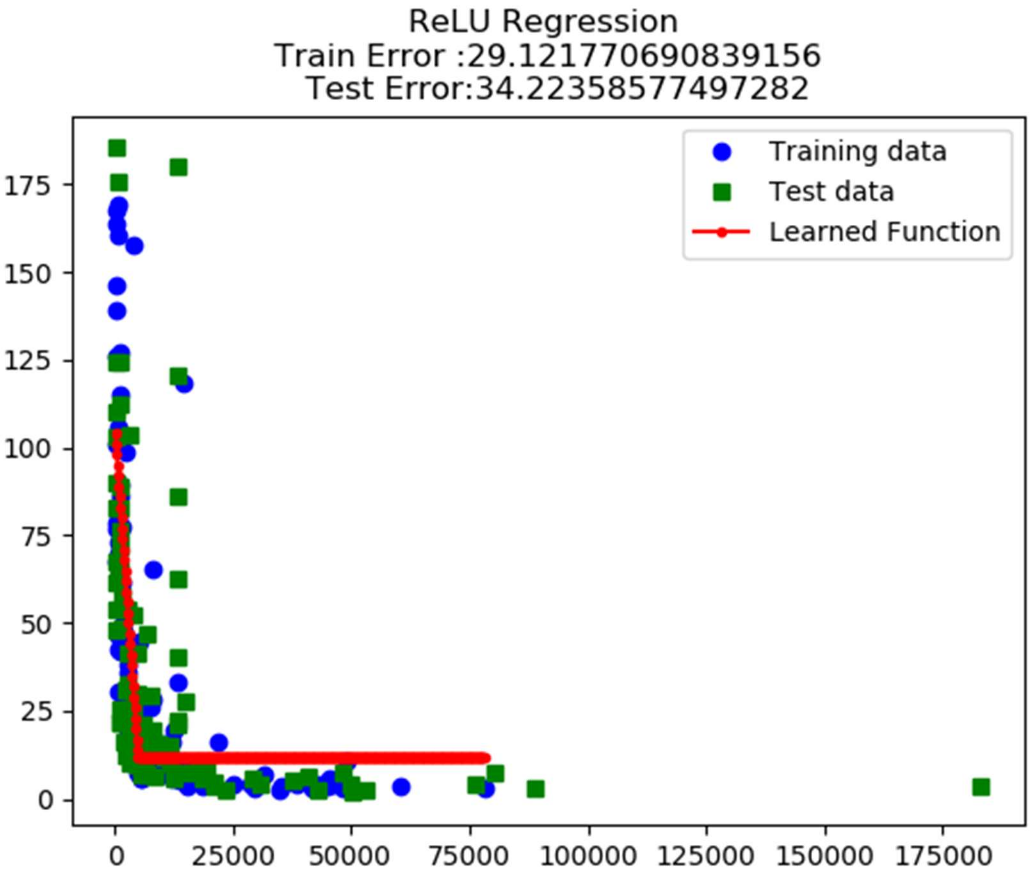
5.2.2.2.


A visualization of a regression estimate using random outputs
Test data points
Training data points
Learned Polynomial
RMS
GNI per capita (US$) 2011
175
150
125
100
75
50
25
0
0
25000
50000
75000
100000
125000
150000
175000


A visualization of a regression estimate using random outputs
Test data points
Training data points
Learned Polynomial
RMS
0
25
50
75
100
125
150
175
30
40
50
60
70
80
90
100
Total adult literacy rate (%) 2007_2011*


A visualization of a regression estimate using random outputs
Test data points
Training data points
Learned Polynomial
RMS
0
25
50
75
100
125
150
175
50
55
60
65
70
75
80
85
Life expectancy at birth (years) 2011

5.3.


ReLU Regression
Train Error :29.121770690839156
Test Error:34.22358577497282
Training data
Test data
Learned Function
175
150
125
100
75
50
25
0
0
25000
50000
75000
100000
125000
150000
175000