CSE 453 - PATTERN RECOGNITION HW1 REPORT

In this homework, I implemented the(vanilla) gradient descent optimization algorithm from scratch. Mean Square Error (MSE) is used as loss function.

Some formulas:

Linear function to estime : $J(\theta)=\theta_0+\theta_1.X$

MSE : $\frac{1}{m}\sum_{i=1}(h_\theta(x^{(i)})-y^{(i)})$

$$Gradient\ of\ \theta_0: \frac{\partial x}{\partial \theta_0}=\frac{\partial}{\partial \theta_0}\frac{1}{m}\sum_{i=1}(\theta_1 x^{(i)}+\theta_0-y^{(i)})^2=\frac{2}{m}\sum_{i=1}(h_\theta(x^{(i)})-y^{(i)})$$

$$Gradient\ of\ \theta_1: \frac{\partial x}{\partial \theta_1}=\frac{\partial}{\partial \theta_1}\frac{1}{m}\sum_{i=1}(\theta_1 x^{(i)}+\theta_0-y^{(i)})^2=\frac{2}{m}\sum_{i=1}(h_\theta(x^{(i)})-y^{(i)}).x^{(i)}$$

m : Number of examples

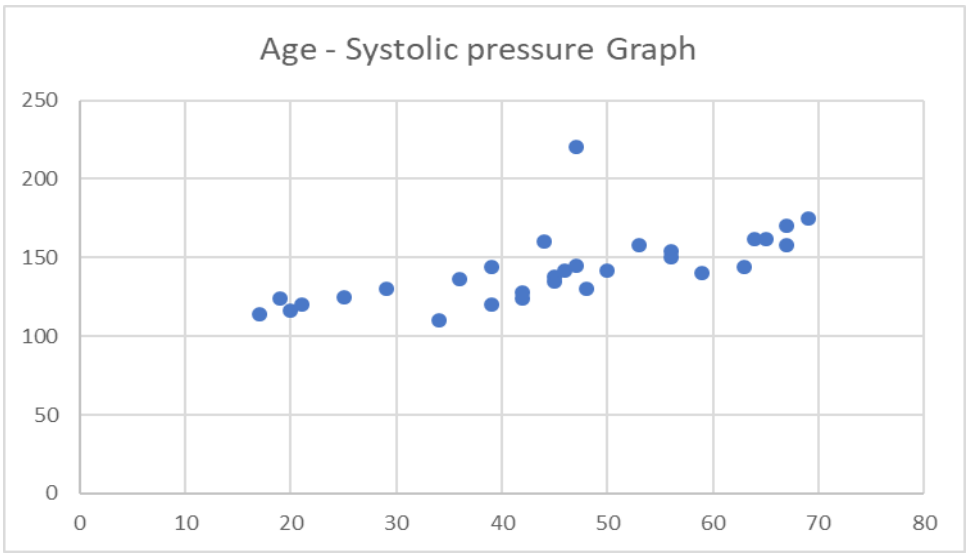
$$\delta_t \leftarrow \alpha\,\delta_{t-1}-\eta\,\nabla_{\theta_{t-1}} f(\theta)$$

$$\theta_t \leftarrow \theta_{t-1}+\delta_t$$

η: learning rate

α: momentum

5 different experiments are carried out to test effecct of parameters to performance of model. 25 of 30 samples are used to calculate regression model and 5 of 30 samples are used to test calculated model. The samples consist of systolic pressure values corresponding to age.

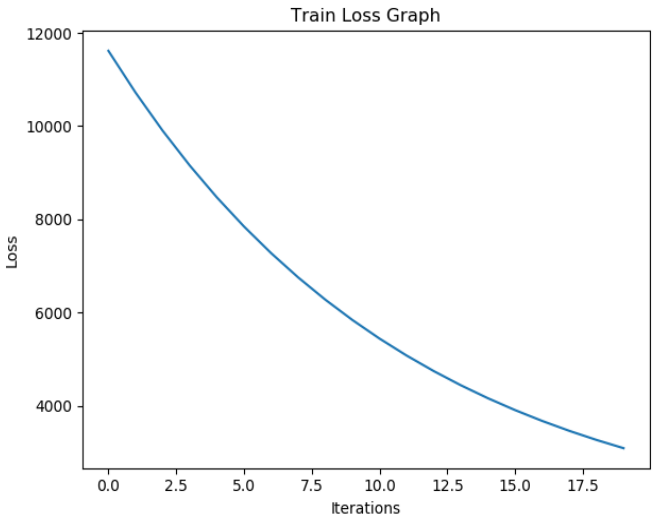


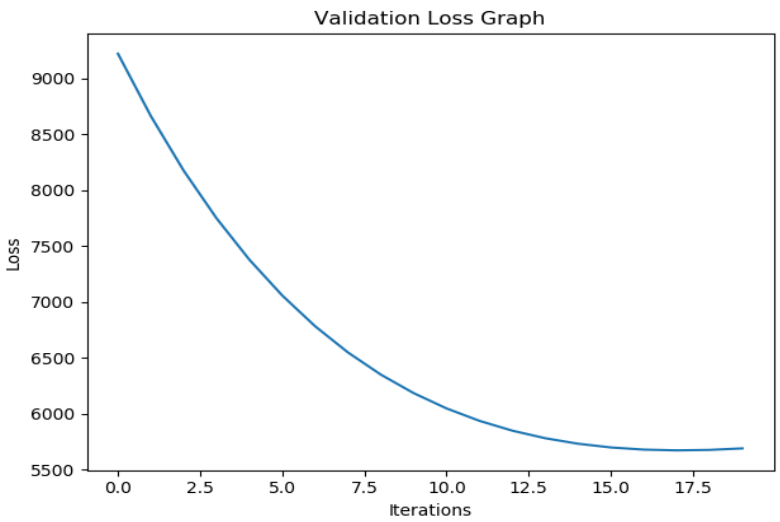
Regression Graph of Data

Experiment 1

In this experiment, performance of model is tested. Initial parameters are randomly selected. Training parameters are as follow.

Number of iterations : 20
Learning rate : 0.00001
Momentum : 0





Train Loss : 3095.8347829174268
Test Loss : 5691.2815735844415

Experiment 2

In this experiment, effect of learning rate on convergence speed and performance are tested.

Learning Rate	Number of iteration to converge	Train Loss	Test Loss
0.5	94	nan	nan
0.1	118	nan	nan
0.05	132	nan	nan
0.01	189	nan	nan
0.005	233	nan	nan
0.001	576	nan	nan
0.0005	3908	nan	nan
0.0001	42407	499.9993887210	5918.1053462184
0.00005	84422	499.9975819434	5918.0970018389
0.00001	422540	499.9996508247	5918.1065567095

Experiments show that, as the learning rate decreases, the convergence speed decreases and to achieve best performance, the model should be further trained in low learning rates.

Experiment 3

In this experiment, effect of learning rate on convergence speed is tested.

Batch Size	Number of iteration to converge
5	130
10	193
15	295
20	262
25	576

Experiment 4

In this experiment, different initialization on convergence speed is tested.

Initial values	Number of iteration to converge
Weight = 0, bias = 0	576
Weight = 1, bias = 1	576
Weight = 2, bias = 2	577
Weight = 3, bias = 3	579
Weight = 4., bias = 04	577
Weight = 5, bias = 5	576
Weight = 6, bias = 6	576
Weight = 7, bias = 7	575
Weight = 8, bias = 8	575

Experiments show that, different initializations don’t effect convergence speed and performance of model.

Experiment 5

In this experiment, effect of momentum on convergence speed and performance.

Learning Rate	Number of iteration to converge	Train Loss	Test Loss
0	576	nan	nan
0.1	595	nan	nan
0.2	617	nan	nan
0.3	642	nan	nan
0.4	671	nan	nan
0.5	705	nan	nan
0.6	748	nan	nan
0.7	802	nan	nan
0.8	873	nan	nan
0.9	974	nan	nan
1	1136	nan	nan

Experiments show that, as the momentu increases, the convergence speed decreases and to achieve best performance, the model should be further trained in high momentum values.