$1 - 5$

standardization

standardization

$$\frac{x - \bar{x}}{\sigma}$$

(정규화 하고자 하는 값 - 데이터의 평균) / 데이터의 표준편차

Without standardization

$$y = x + 1$$

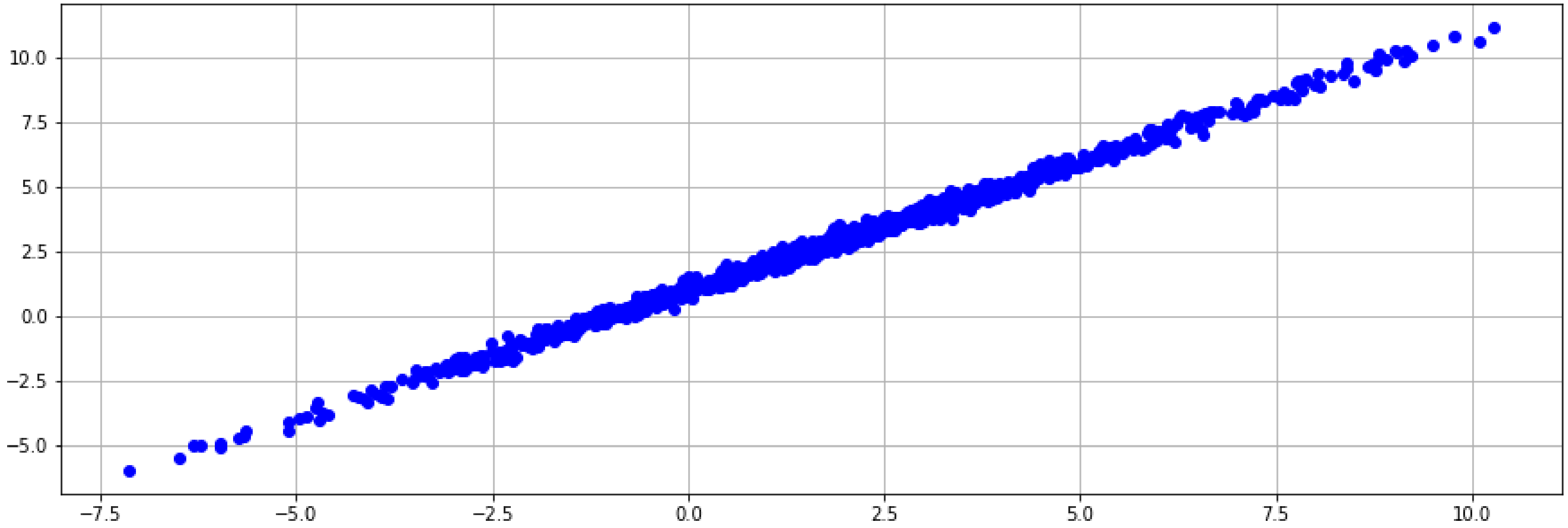
```
n_data = 1000
data_mean, data_std = 2., 3.

##### Your Code(Dataset Generation/Start) #####
x_data = np.random.normal(data_mean,data_std,size = n_data)
y_data = 0.2*np.random.normal(loc = 0, scale = 1., size = n_data) + x_data +1
```

데이터의 개수 : 1000

mean(평균) : 2

std (표준편차) : 3

Data



Bias and Weight

```
ROI = [-1, 3] # Region of Interest
n_point = 30 # number of points
```

Bias 고정   Weight 범위

```
w_range = np.linspace(start = ROI[0] ,stop = ROI[1] ,num = n_point ).reshape(-1,1)
```

Weight 고정 Bias 범위

```
b_range = np.linspace(start = ROI[0] ,stop = ROI[1] ,num = n_point ).reshape(-1,1)
```

Prediction

$$y = \theta_1 * x + \theta_0$$

```
theta1 = 1
theta0 = 1
pred1 =  w_range*x_data+ theta0 # bias가 고정된 prediction
pred2 = theta1*x_data + b_range # weight가 고정된 prediction
```

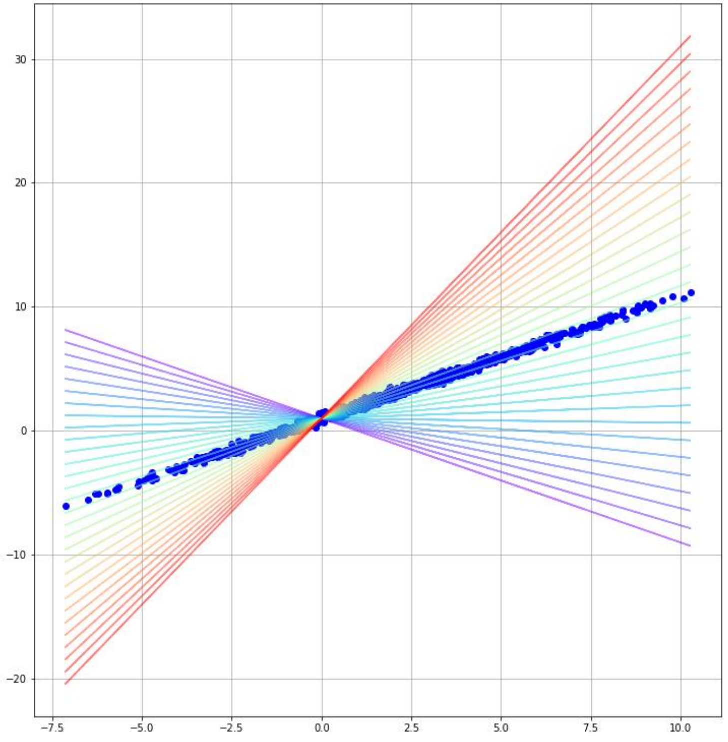
b_range	float64	(30, 1)
w_range	float64	(30, 1)

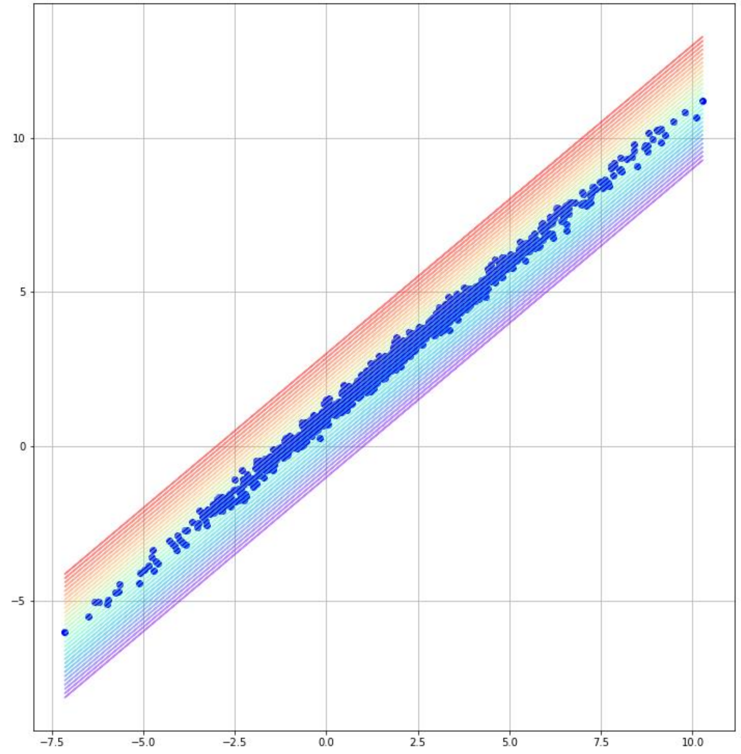
x_data	float64	(1000,)

Prediction

$$y = \theta_1 * x + \theta_0$$

pred1	float64	(30, 1000)
pred2	float64	(30, 1000)





Cost function

$$cost1 = \frac{1}{N}\sum_{i=1}^{N}(y_{data} - pred1)^2$$

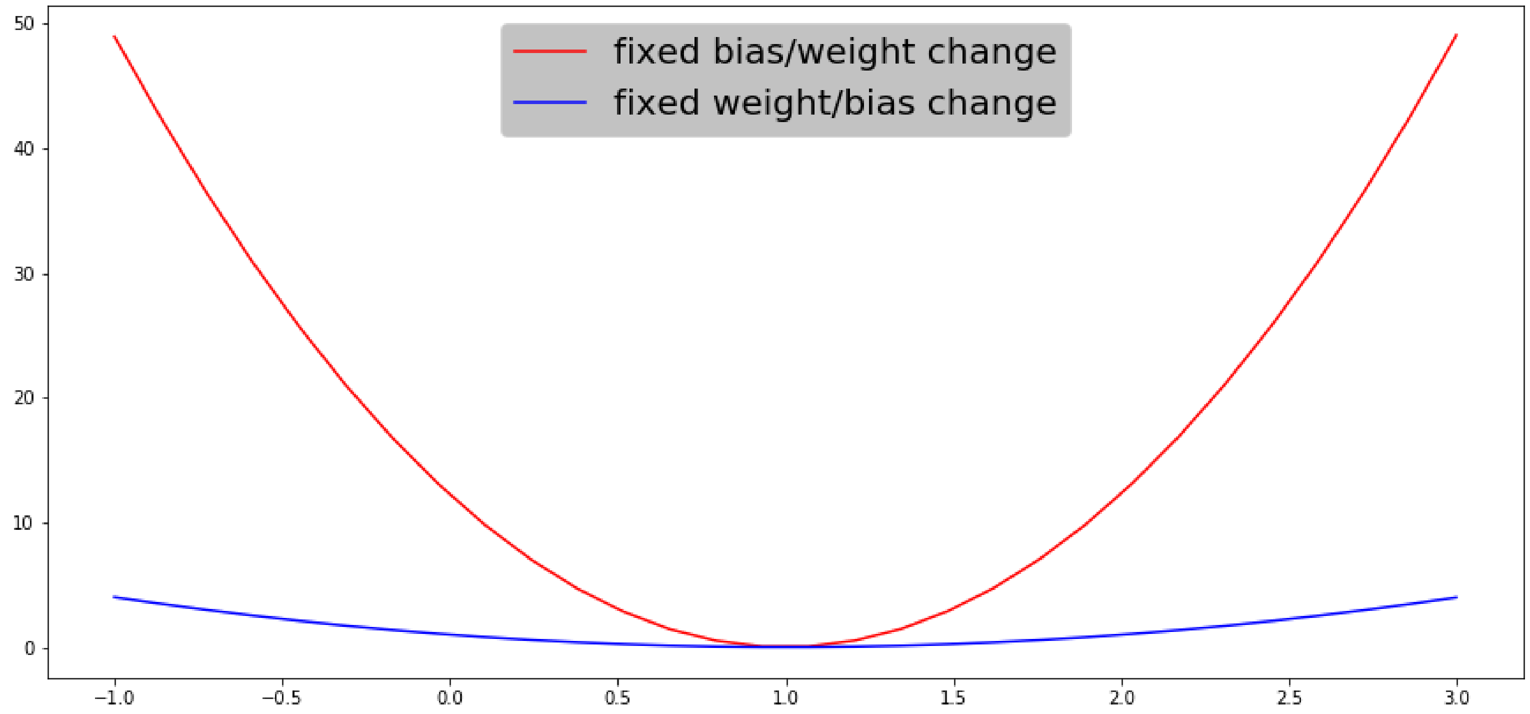
$$cost2 = \frac{1}{N}\sum_{i=1}^{N}(y_{data} - pred2)^2$$

```
cost1 = np.mean(np.power(pred1 - y_data.reshape(1,-1), 2), axis = 1)
cost2 = np.mean(np.power(pred2 - y_data.reshape(1,-1), 2), axis = 1)
```

result

mean(평균) : 2

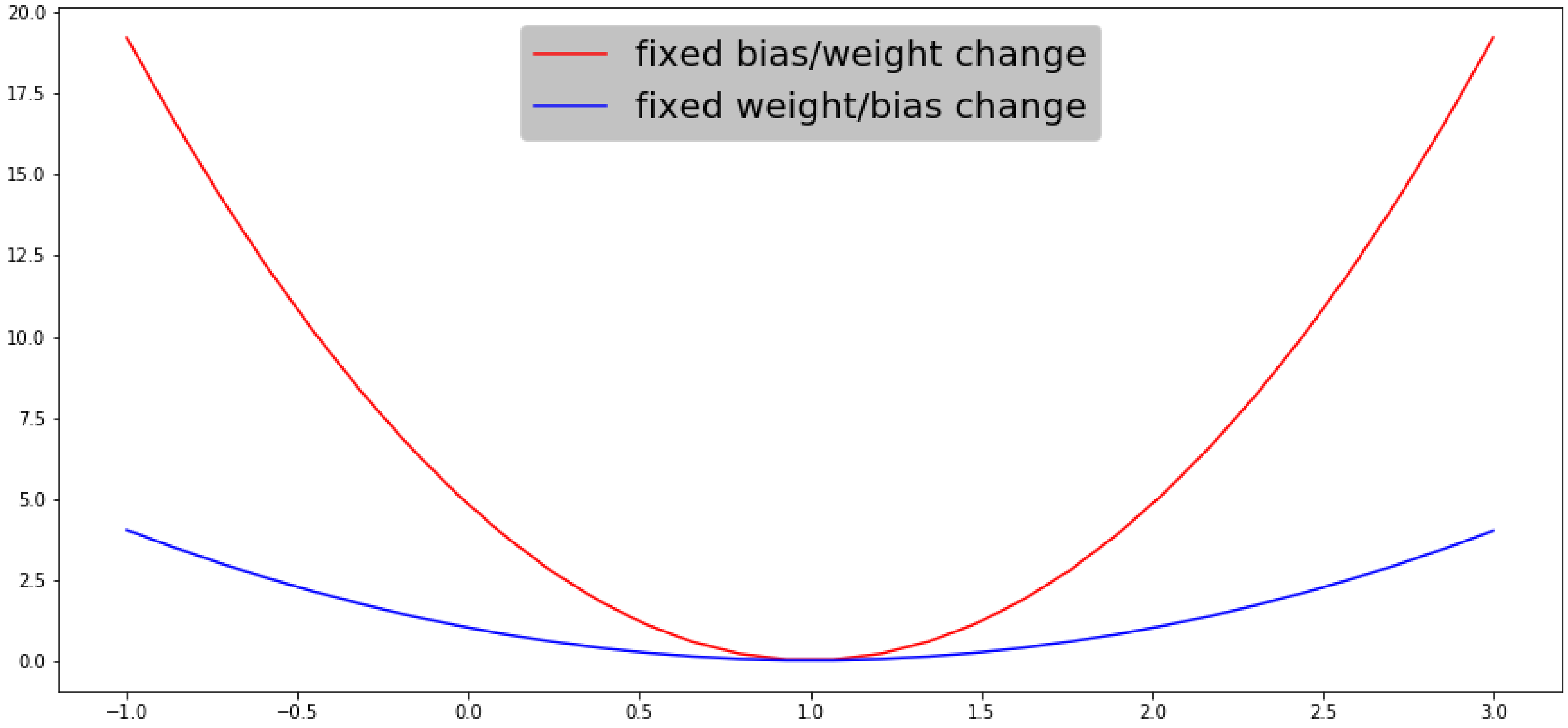
std (표준편차) : 3



result

mean(평균) : 2

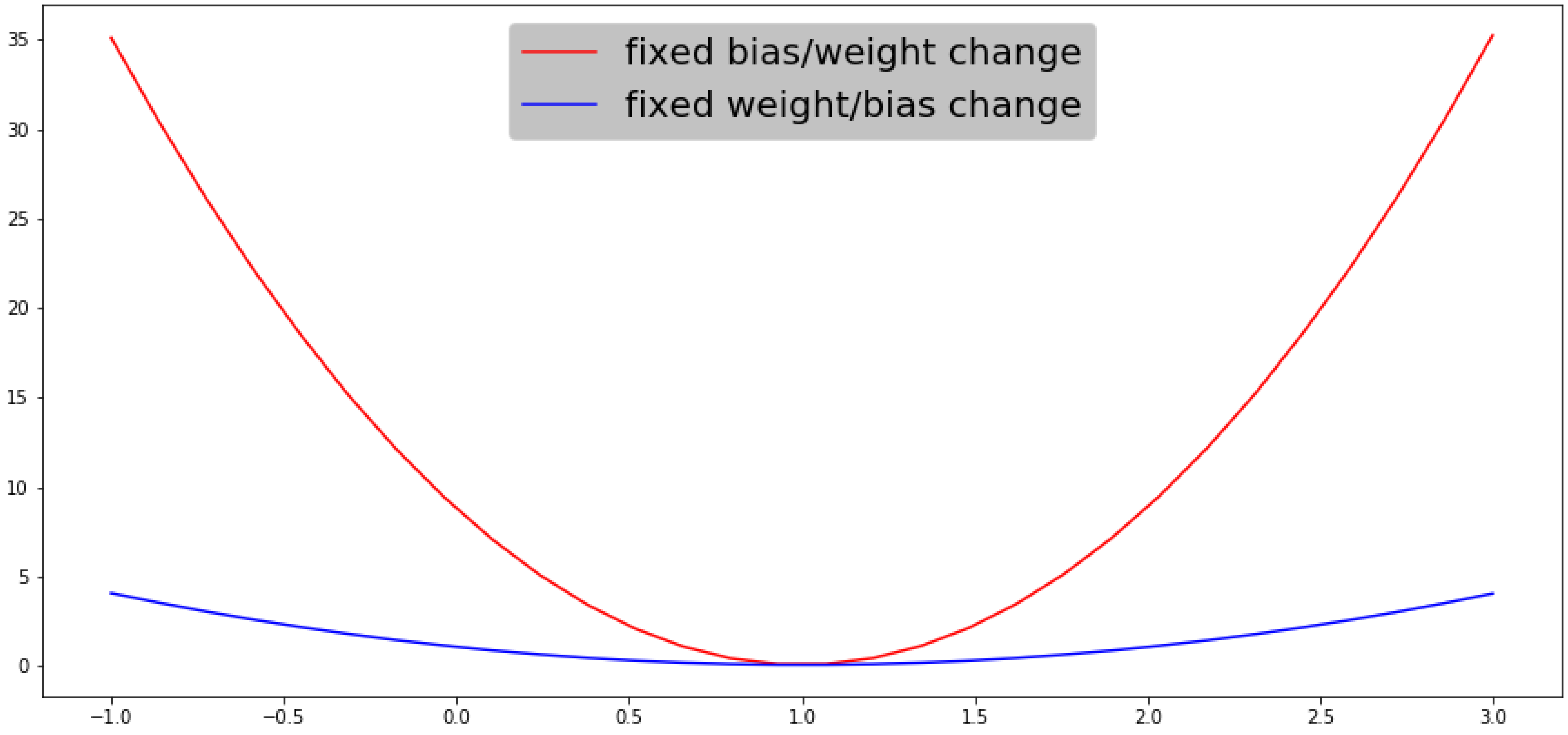
std (표준편차) : 1



result

mean(평균) : 0

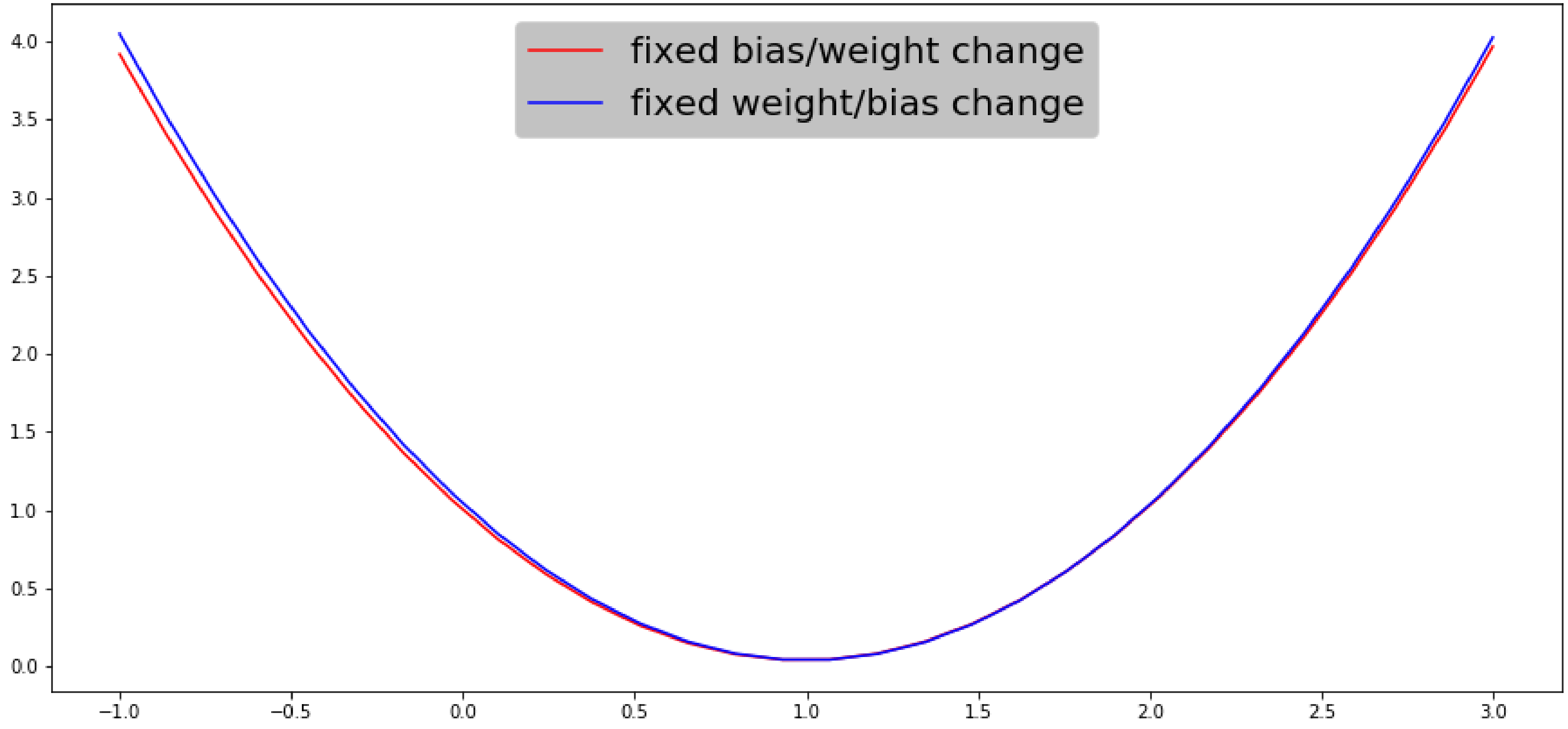
std (표준편차) : 3



result

mean(평균) : 0

std (표준편차) : 1



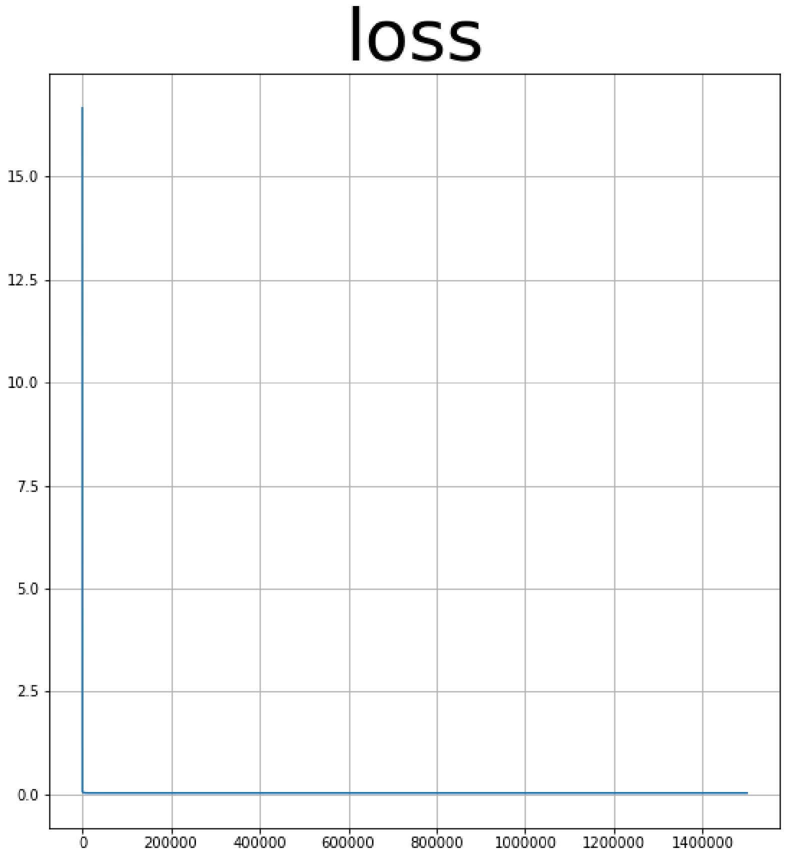
Q1)x data의 mean, std이 learning에 미치는 영향에 대해서 분석하시오

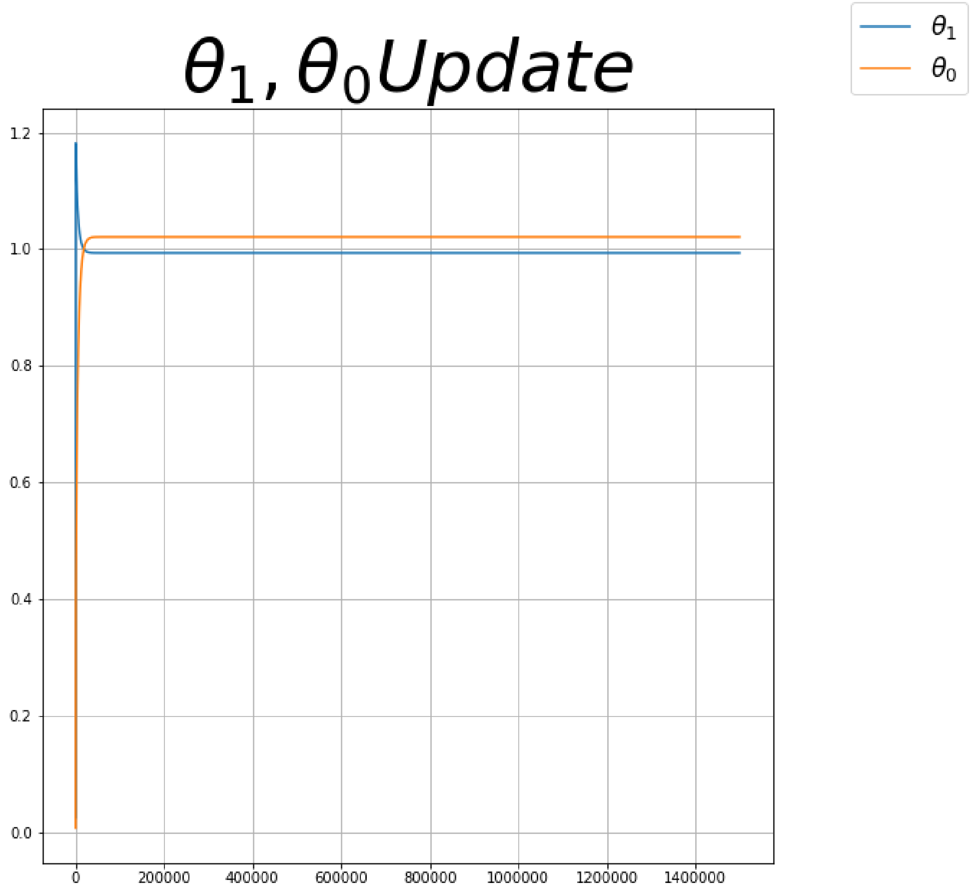
Q2)위와 dataset이 주어졌을 때, mean, std의 허용범위를 찾아보고 그 근거를 제시하시오

Q3)mean, std을 각각 0, 1로 조절하는 과정 중 어떤 것이 learning에 더 영향을 많이 미치는지 분석하시오

mean(평균) : 3

std (표준편차) : 1





mean(평균) : 0

std (표준편차) : 32