ASSIGNMENT 1

180050023
Bhaskar Gupta

1.1

$$mse(w,b) = \frac{1}{2N}\sum_{i=1}^{N}(wx_i + b - y_i)^2$$

$$\frac{\delta\, mse(w,b)}{\delta b} = \frac{1}{N}\sum_{i=1}^{N}(wx_i + b - y_i)$$

$$\frac{\delta\, mse(w,b)}{\delta w} = \frac{1}{N}\left(\sum_{i=1}^{N}(wx_i + b - y_i)x_i\right)$$

$$\nabla mse(w,b) = \begin{bmatrix} \frac{\delta}{\delta w} mse(w,b) \\ \frac{\delta}{\delta b} mse(w,b) \end{bmatrix}$$

1.2 (d) Resulting line in the figure is the Best Fit, not necessarily passing through all points

2.1 (a) $\hat{Y} = XW$

(b) $mse(W) = \frac{1}{2N}\sum_{i=1}^{N}(x_i w_i - y_i)^2 = \cancel{\frac{1}{2N}(w'}$

$$mse(W) = \frac{1}{2N}(XW - Y)^T(XW - Y)$$

$$= \frac{1}{2N}(W^T X^T X W - W^T X^T Y - Y^T X W + Y^T Y)$$

$(W^T X^T Y = (Y^T X W)^T)$

as $Y^T X W$ is 1×1

$$\text{mse}(W) = \frac{1}{2N}\left(Y^T Y - 2W^T X^T Y + W^T (X^T X) W \right)$$

Gradient of mse (W) is as follows

$$\nabla \text{mse}(W) = \frac{1}{2N}\left(\nabla Y^T Y - 2\nabla W^T X^T Y + \nabla W^T (X^T X) W \right)$$

$$= \frac{1}{2N}\left(0 - 2X^T Y + 2X^T X W \right)$$

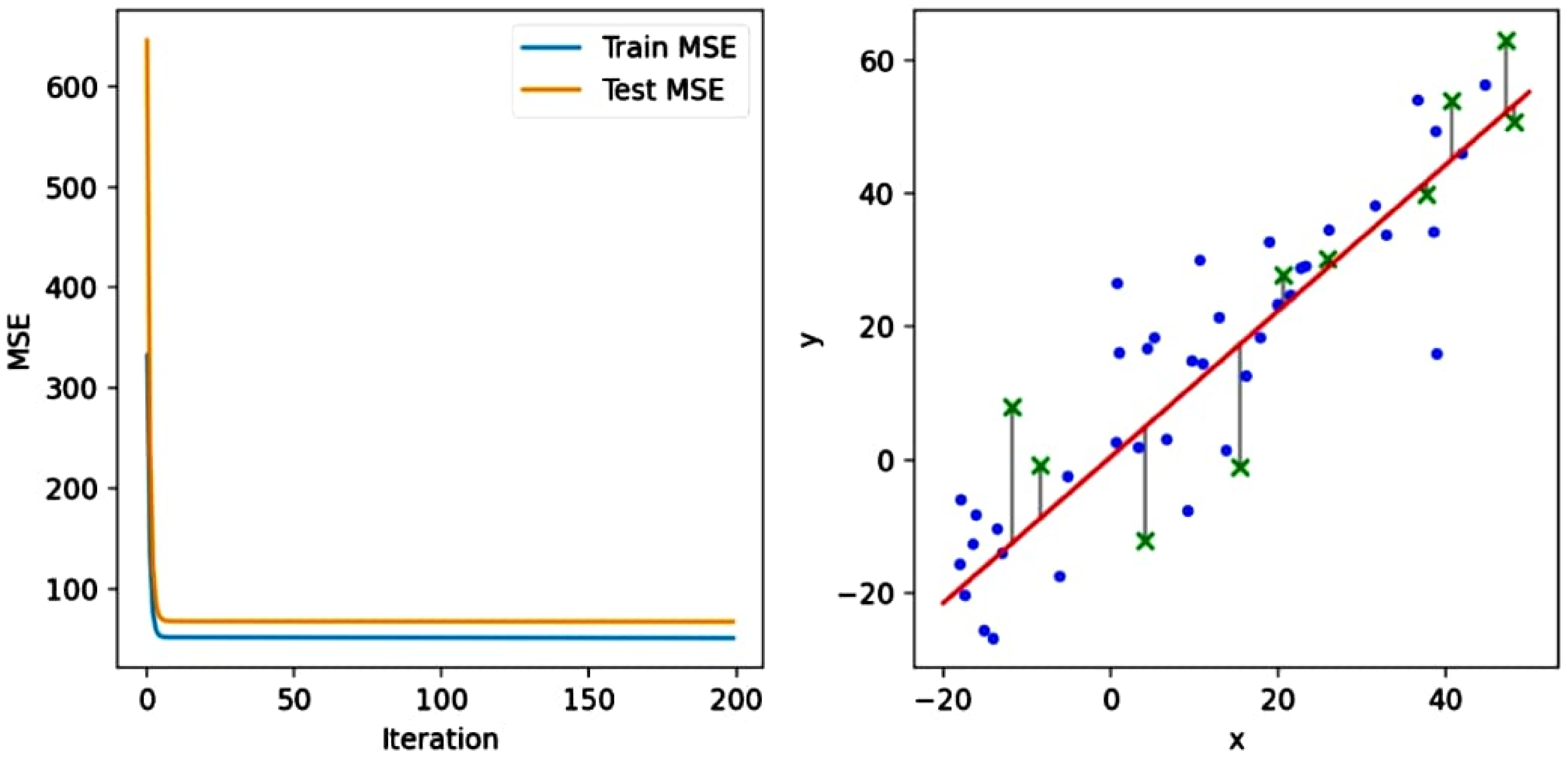
$$\nabla \text{mse}(W) = \frac{1}{N}\left(X^T X W - X^T Y \right)$$

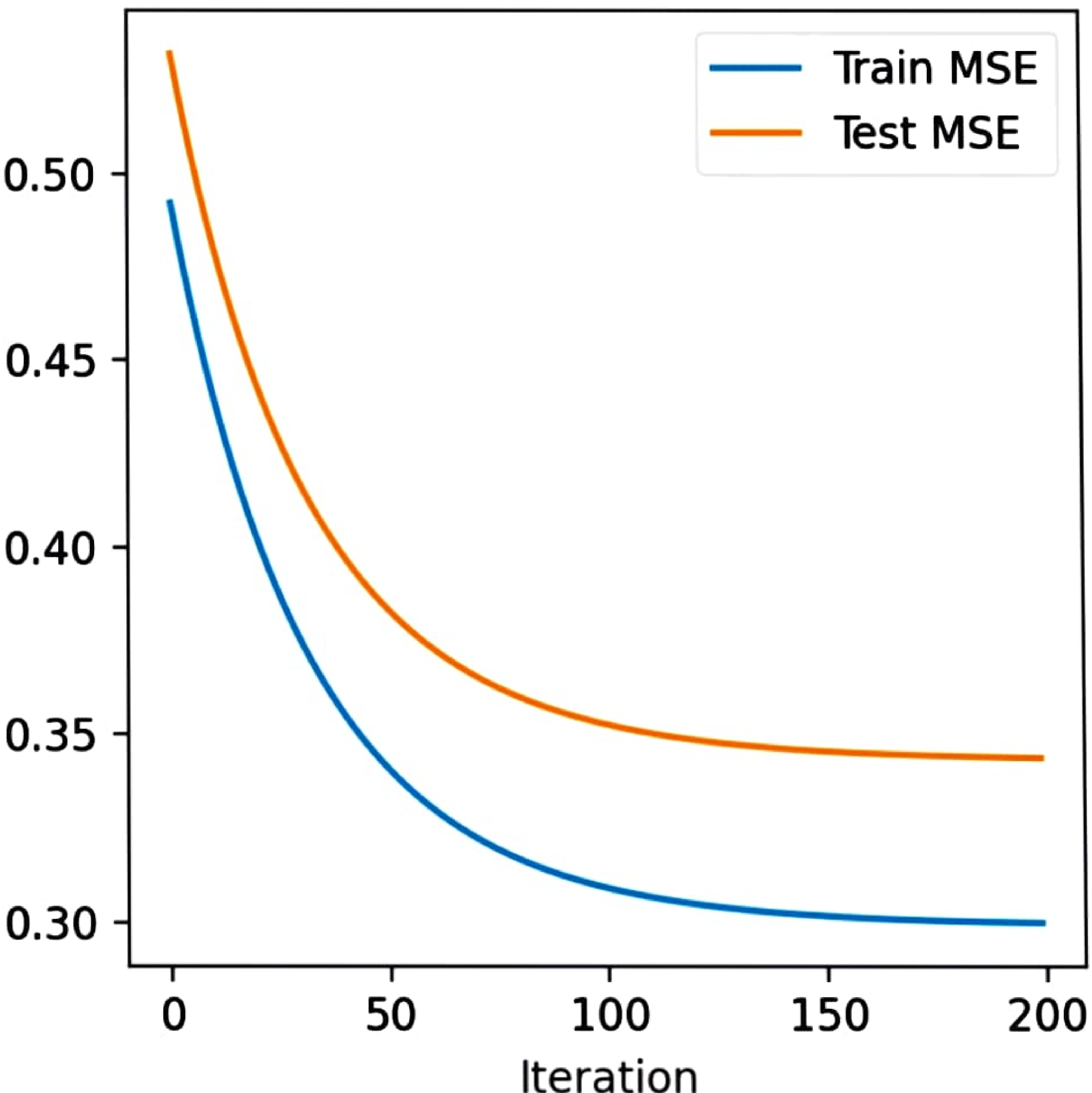
$$\boxed{\nabla \text{mse}(W) = \frac{1}{N} X^T (XW - Y)}$$

2.1 (c) $\quad \text{mse}(W) = \frac{1}{2N}\left(\sum_{i=1}^{N} (XW - Y)^2 \right) + \lambda W^T W$

$$\text{mse}(W) = \frac{1}{2N} (XW - Y)^T (XW - Y) + \lambda W^T W$$

$$\boxed{\nabla \text{mse}(W) = \frac{1}{N} X^T (XW - Y) + 2\lambda W}$$


Train MSE
Test MSE
MSE
Iteration
600
500
400
300
200
100
0
50
100
150
200
y
x
60
40
20
0
−20
−20
0
20
40



Train MSE
Test MSE
0.50
0.45
0.40
0.35
0.30
0
50
100
150
200
Iteration

3.1 $$E(W) = \frac{1}{2N} \sum_{i=1}^{N} r_i^2 (y_i - W^T x_i)^2$$
$$= \frac{1}{2N} \sum_{i=1}^{N} (r_i y_i - W^T r_i x_i)^2$$

let R be a diagonal matrix

with $R[i,i] = r_i$ for $i = 1:n$

now X can be $\rightarrow RX$

Y can be $\rightarrow RY$

Closed form for linear regression

is $w^* = (X^T X)^{-1} X^T Y$

Substituting the weighted points

we get $\boxed{w^* = (X^T R^2 X)^{-1} X^T R^2 Y}$

4.1 1st and 3rd column are linearly dependent hence $X^T X$ was not ~~full rank so i be~~ invertible as X is not full rank

So i just removed the first column as it is redundant

4.2 When X has linearly dependent columns i.e it is not full rank X^TX is a singular matrix which lead to the failure of above regression but gradient descent will still converge to a best fit solution as the cost function is convex up hence we reach the global minima regardless as we use gradient descent