Simple Linear Regression

Linear Regression

- Linear Regression is a prediction/estimation method
- It requires you to estimate the parameters from the training data

Linear Regression

- Estimate the relationship between a **dependent variable** (usually called Y) and an **independent variable** (usually called X)

$$Y = B0 + B1 . X$$

Y – Output variable to be predicted

X – input variable given

B0 and B1 – parameters (or coefficients) to be estimated from training data.

Linear Regression

- **Training :** Determine parameters B0 and B1 from the training data

- **Testing** : Use the estimated parameters B0 and B1 in the regression equation to predict the value of Y for given test data X.

Linear Regression

- Estimated regression line based on sample data is given as:

$$\hat{Y} = B_0 + B_1 X$$

- The method of least squares chooses the values for B_0, and B_1 to minimize the sum of squared errors:

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \left(Y - B_0 - B_1 X\right)^2$$

Linear Regression

In order to get the parameters for best fit lines, we need to:

1. Partially differentiate the SSE w.r.t B_0 and equate to 0. Solve this equation for value of B_0
2. Similarly, partially differentiate the SSE w.r.t B_1 and equate to 0. Solve this equation for the value of B_1

Determine B0 and B1

$$B1 = \frac{\sum_{i=1}^{n}((x_i - mean(x)) \times (y_i - mean(y)))}{\sum_{i=1}^{n}(x_i - mean(x))^2}$$

$$B0 = mean(y) - B1 \times mean(x)$$

Example

- Example

  The weekly advertising expenditure (x) and weekly sales (y) are presented in the following table.

y	x
1250	41
1380	54
1425	63
1425	54
1450	48
1300	46
1400	62
1510	61
1575	64
1650	71

Example

– From previous table we have:

$$n = 10 \qquad \sum x = 564 \qquad \sum x^2 = 32604$$
$$\sum y = 14365 \qquad \sum xy = 818755$$

– The least squares estimates of the regression coefficients are:

$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{10(818755) - (564)(14365)}{10(32604) - (564)^2} = 10.8$$

$$b_0 = 1436.5 - 10.8(56.4) = 828$$

Example

- The estimated regression function is:

$$\hat{y} = 828 + 10.8x$$

$$\text{Sales} = 828 + 10.8\ \text{Expenditut}\ \text{e}$$

- This means that if the weekly advertising expenditure is increased by $1 we would expect the weekly sales to increase by $10.8.

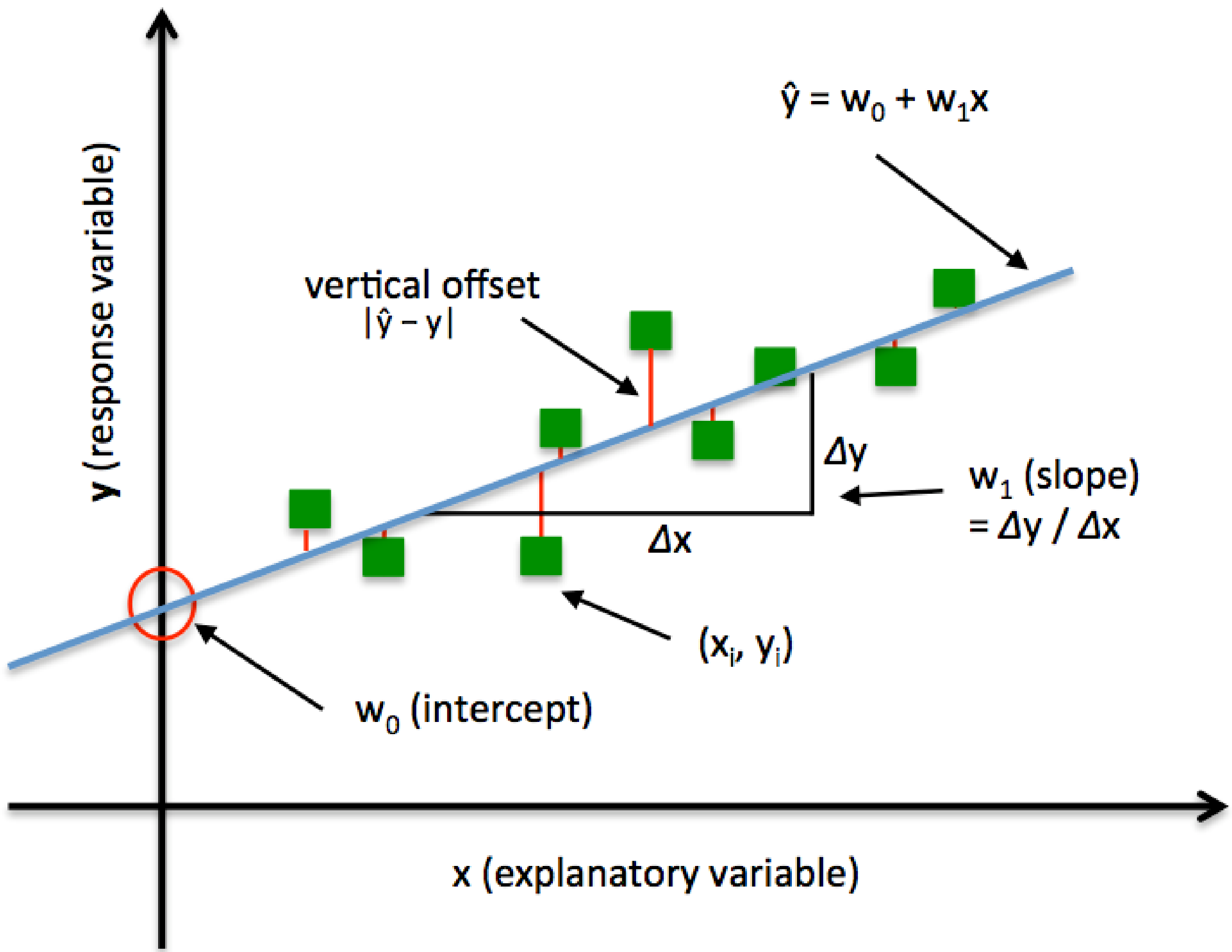
Example

- Fitted values for the sample data are obtained by substituting the x value into the estimated regression function.
- For example if the advertising expenditure is $50, then the estimated Sales is:

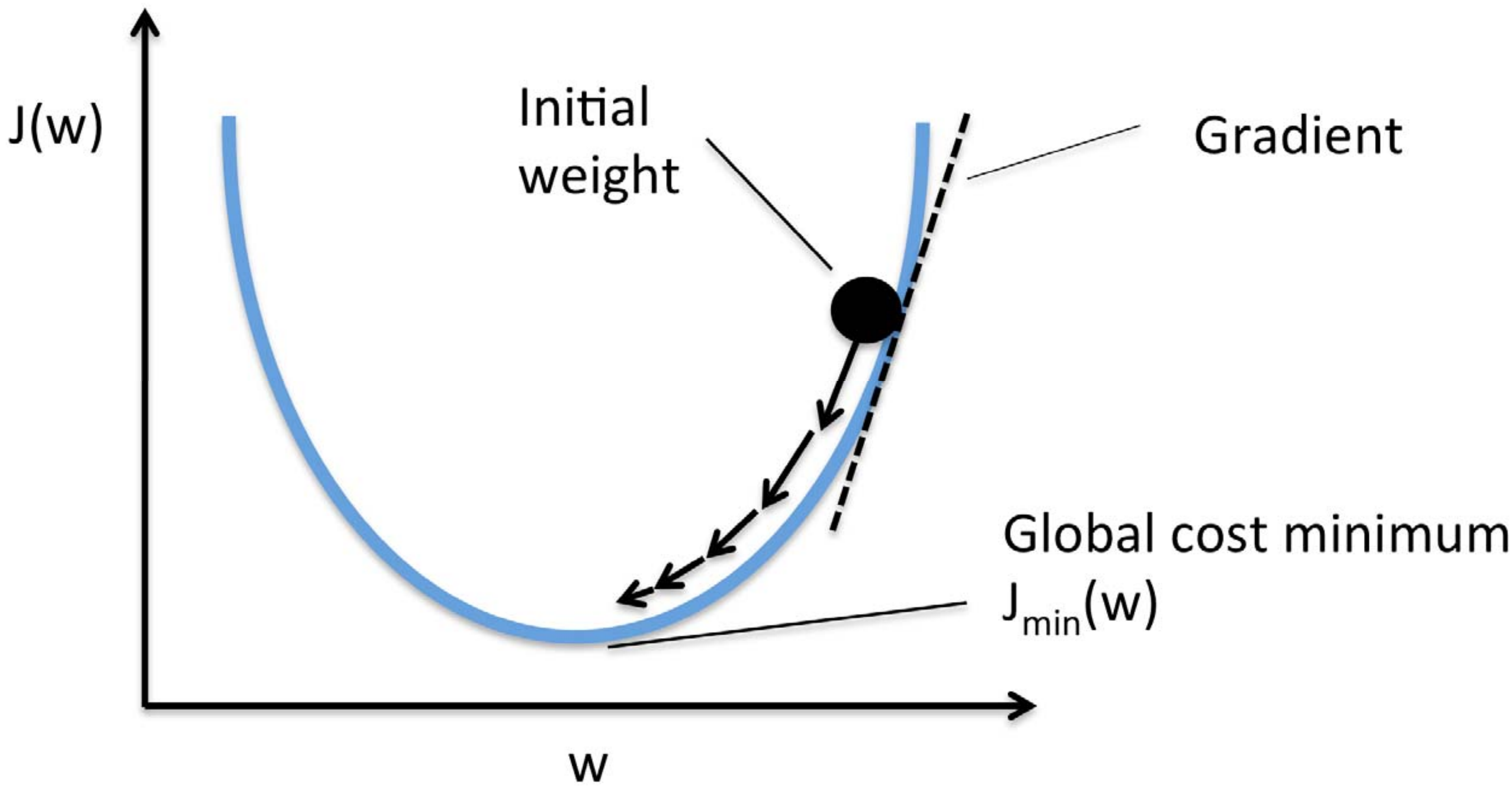
$$Sales = 828 + 10.8(50) = 1368$$

- This is called the point estimate (forecast) of the mean response (sales).

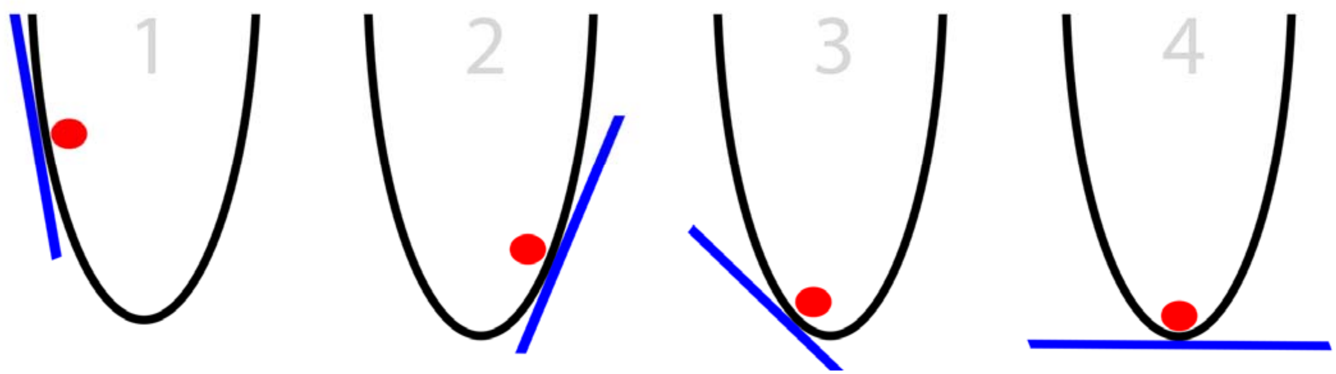
Gradient Decent



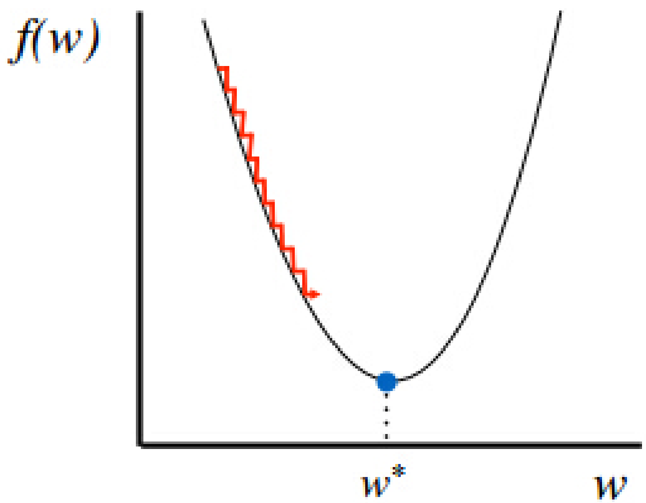
Gradient Decent



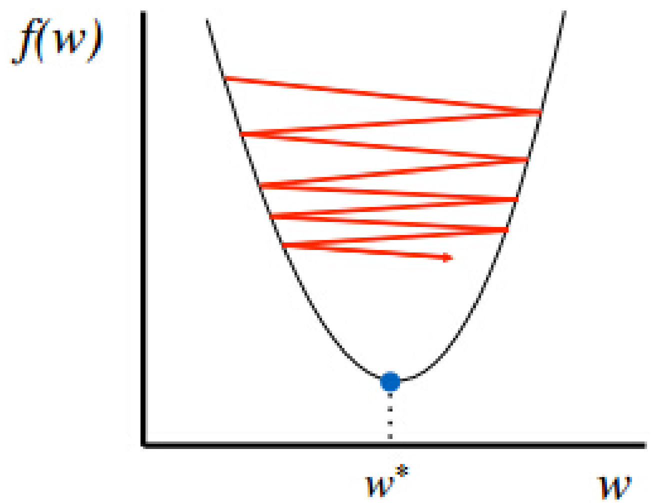
Gradient Decent



Gradient Decent

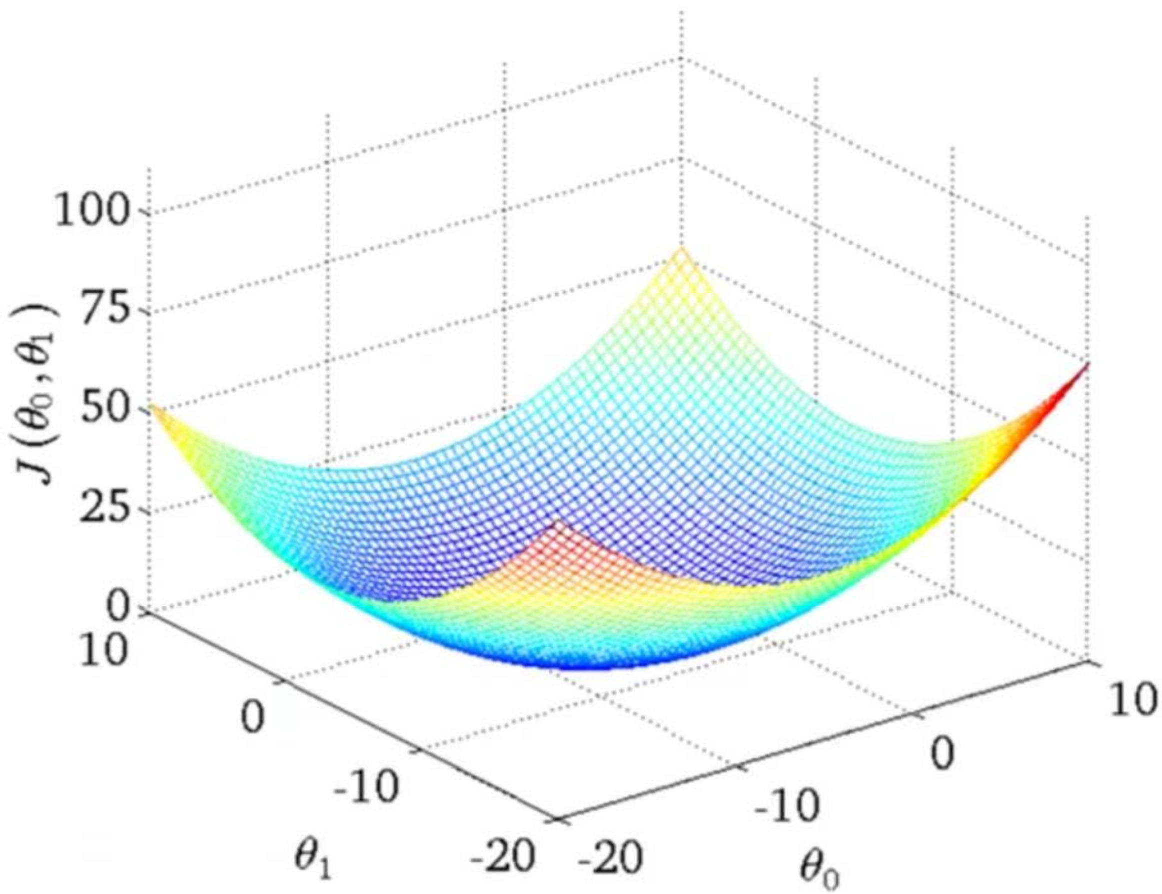


Too small: converge very slowly

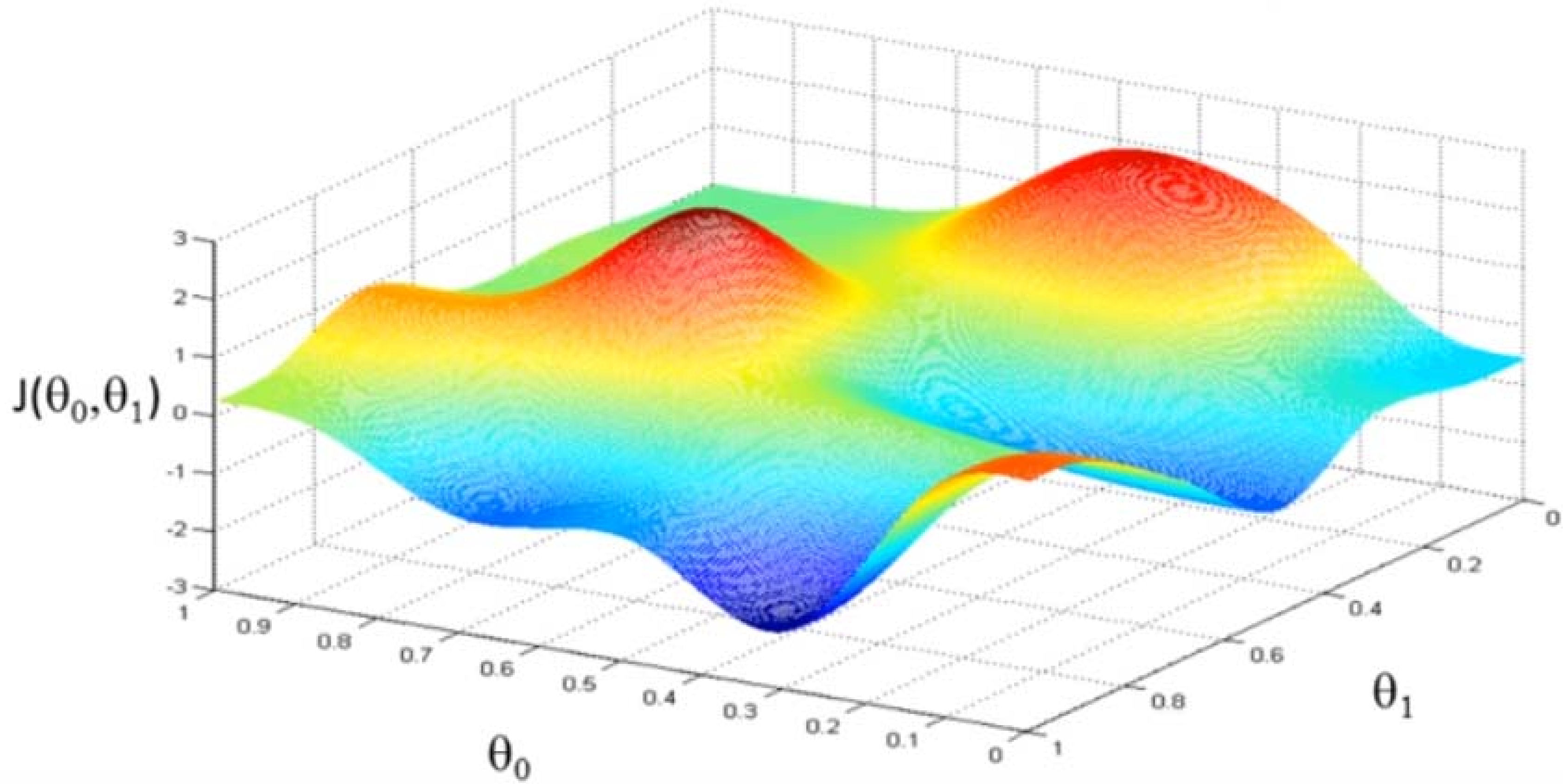


Too big: overshoot and even diverge

Gradient Decent



Gradient Decent



Gradient Decent