MACHINE LEARNING

Shoaib Farooq

Department of Computer Science

Instructor Information

Instructor	Shoaib Farooq	E-mail	m.shoaib1050@gmail.com
GitHub	https://github.com/SHOAIB1050		
YouTube	https://www.youtube.com/c/pakithub		
LinkedIn	https://www.linkedin.com/in/shoaib-farooq-b5b190105/		

Let's Start .....

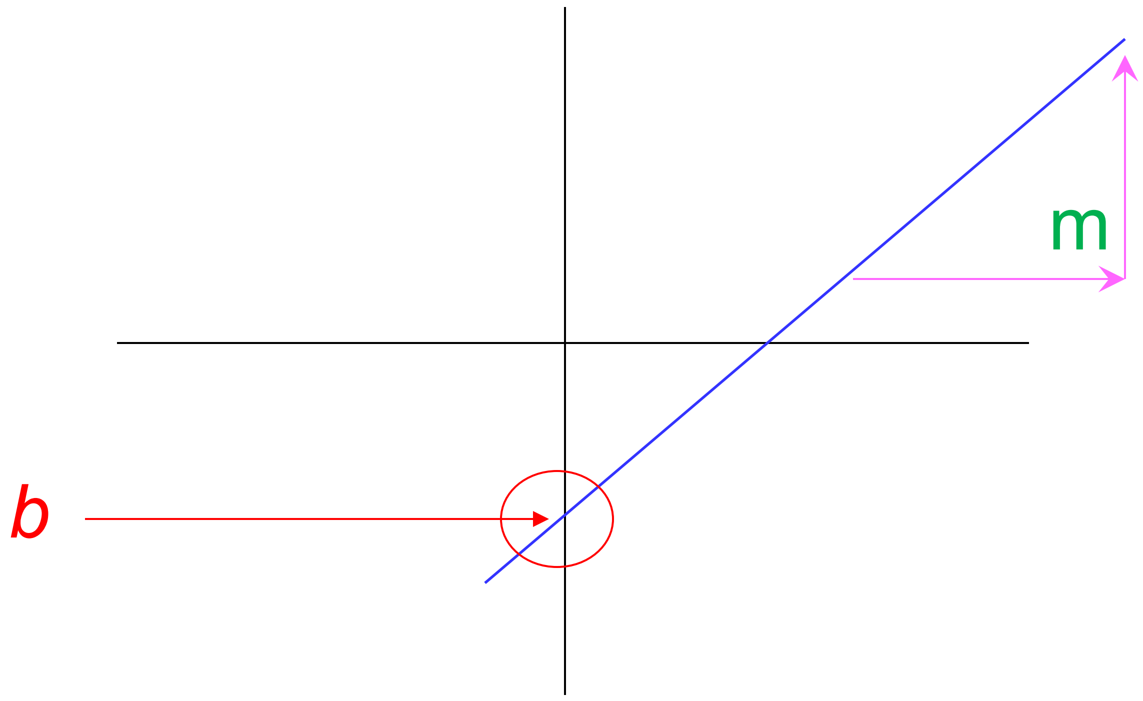
Lecture #6

GOALS

- **This Lecture Will Cover:**
- **Linear Regression**

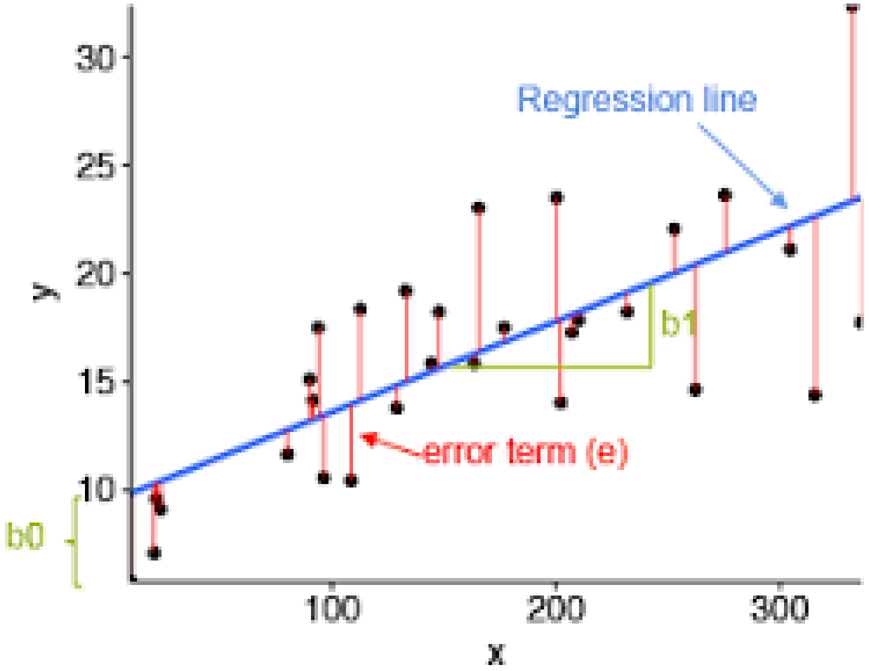
WHAT IS "LINEAR"?

- Remember this:
- $Y = mX + b$?



PREDICTION

If you know something about X, this knowledge helps you predict something about Y.

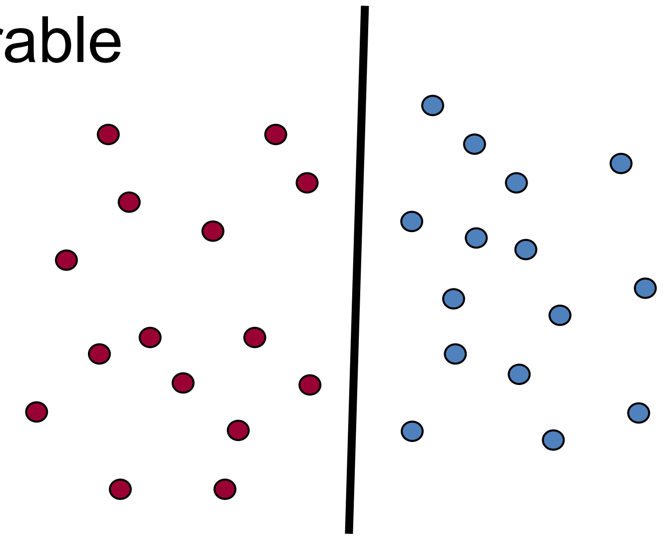


LINEAR MODELS

An assumption is *linear separability*:

- in 2 dimensions, can separate classes by a line
- in higher dimensions, need hyperplanes

A *linear model* is a model that assumes the data is linearly separable



LINEAR REGRESSION

Linear regression is one of the simplest and most widely used algorithms in machine learning, primarily for predictive modeling in regression problems. The goal of linear regression is to model the relationship between a dependent variable (often called the target or output) and one or more independent variables (features or predictors) by fitting a linear equation to observed data.

LINEAR REGRESSION

- The goal of simple linear regression (univariate) model is to finds the relation between two variable.
  - A single feature (variable x) and a continuous valued response (target variable y).
  - X is called independent variable (predictor)
  - Y is called the dependent (target or response) variable.

$$y = w_0 + w_1 x$$

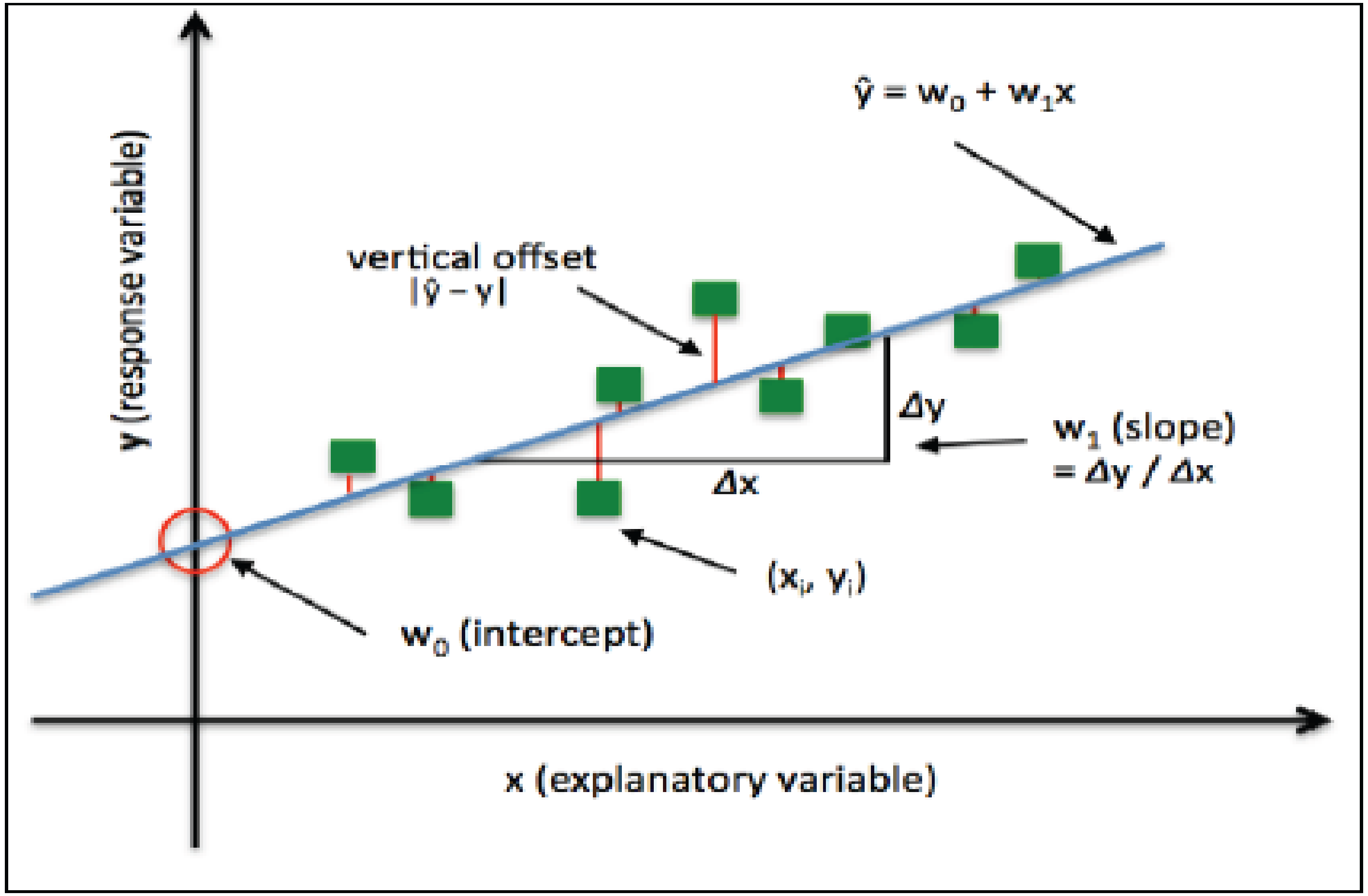
LINEAR REGRESSION

- The weight w_0 represents the y axis intercepts and w_1 is the coefficient of the feature (x variable).
  - w_0 and w_1 are unknown
- The goal of linear regression is to learn the weights of the linear equation
  - Describe the relationship between the **x** and **y**
- Then this relation can be used to predict the responses of new data

LINEAR REGRESSION

- Linear regression can be understood as finding the best-fitting straight line through the sample points.
- This best-fitting line is also called the **regression line.**
- The distance between the regression line to the sample points are the so-called **offsets** or **residuals**
  - The errors of our prediction.

LINEAR REGRESSION



LINEAR MODELS IN GENERAL

- For leaner model:

$$y = w_0 + w_1 x$$

- These are the parameters we want to learn
- Need to define a criteria to optimize these parameters of the model
  - cost function (objective)
  - Minimize the cost function

LINEAR MODELS IN GENERAL

Simple Linear Regression:

- Deals with only one independent variable (feature).
- The model can be expressed as:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Where:

- y is the dependent variable.
- x is the independent variable.
- β_0 is the y-intercept.
- β_1 is the slope (the weight of the feature).
- ϵ is the error term.

LINEAR MODELS IN GENERAL

Multiple Linear Regression:

- Involves more than one independent variable.
- The model can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Where:

- $x_1, x_2, \ldots, x_n$ are the independent variables (features).
- $\beta_1, \beta_2, \ldots, \beta_n$ are the weights (coefficients) associated with each feature.

STEPS TO PERFORM LINEAR REGRESSION

- **Data Preparation:**

Ensure that the data is clean and pre-processed (e.g., handle missing values, scale features if necessary).

- **Fit the Model:**

The algorithm tries to find the best-fitting line by minimizing the residual sum of squares (RSS), which is the difference between the observed data points and the predicted values.

- **Training the Model:**

The linear regression model uses techniques such as Ordinary Least Squares (OLS) to compute the optimal values of the coefficients

- **Prediction:**

Once trained, the model can be used to predict values for unseen data by applying the learned coefficients.

LINEAR REGRESSION EXAMPLE

x_i (Week)	y_j (Sales in Thousands)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

LINEAR REGRESSION EXAMPLE

- Linear regression equation is given by
- $y = a_0 + a_1 * x + e$
- *where*
- $a_1 = \frac{(\overline{xy}) - (\bar{x})(\bar{y})}{\overline{x^2} - \bar{x}^2}$
- $a_0 = \bar{y} - a_1 * \bar{x}$

x_i (*Week*)	y_j (*Sales in Thousands*)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

LINEAR REGRESSION EXAMPLE

	x_i (*Week*)	y_j (*Sales in Thousands*)	x_i^2	$x_i * y_j$
	1	1.2	1	1.2
	2	1.8	4	3.6
	3	2.6	9	7.8
	4	3.2	16	12.8
	5	3.8	25	19
Sum	15	12.6	55	44.4
Average	$\bar{x} = 3$	$\bar{y} = 2.52$	$\overline{x^2} = 11$	$\overline{xy} = 8.88$

LINEAR REGRESSION EXAMPLE

- $a_1 = \frac{(\overline{xy})-(\bar{x})(\bar{y})}{\overline{x^2}-\bar{x}^2} = \frac{8.88-3*2.52}{11-3^2} = 0.66$

- $a_0 = \bar{y} - a_1 * \bar{x} = 2.52 - 0.66 * 3 = 0.54$

LINEAR REGRESSION EXAMPLE

- Regression equation is
- $y = a_0 + a_1 * x$
- $y = 0.54 + 0.66 * x$
- The predicted 7th week sale (when x = 7) is,
- y = 0.54 + 0.66 x 7 = 5.16

Thank You ☺