Regression

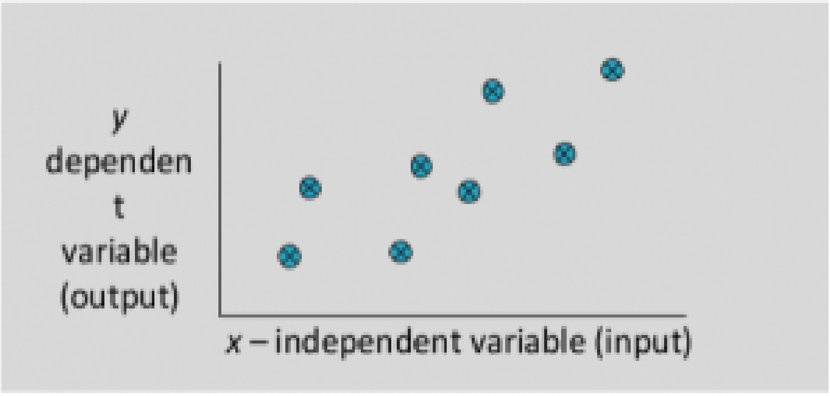
Dr. Mrunal Rane

Syllabus

- Linear Regression
- Gradient Descent
- Lasso and Ridge Regression
- Polynomial Regression
- Logistic Regression
- Maximum Likelihood Function

Regression

- Is Supervised or Unsupervised?
- What is the basic requirement of Supervised learning?
- What is Regression?
  - Regression is a supervised machine learning technique which is used to predict continuous values.
  - Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables.
  - It can be utilized to assess the stren variables and for modeling the future



Which of the following is a regression task?

1

Predict age of a person

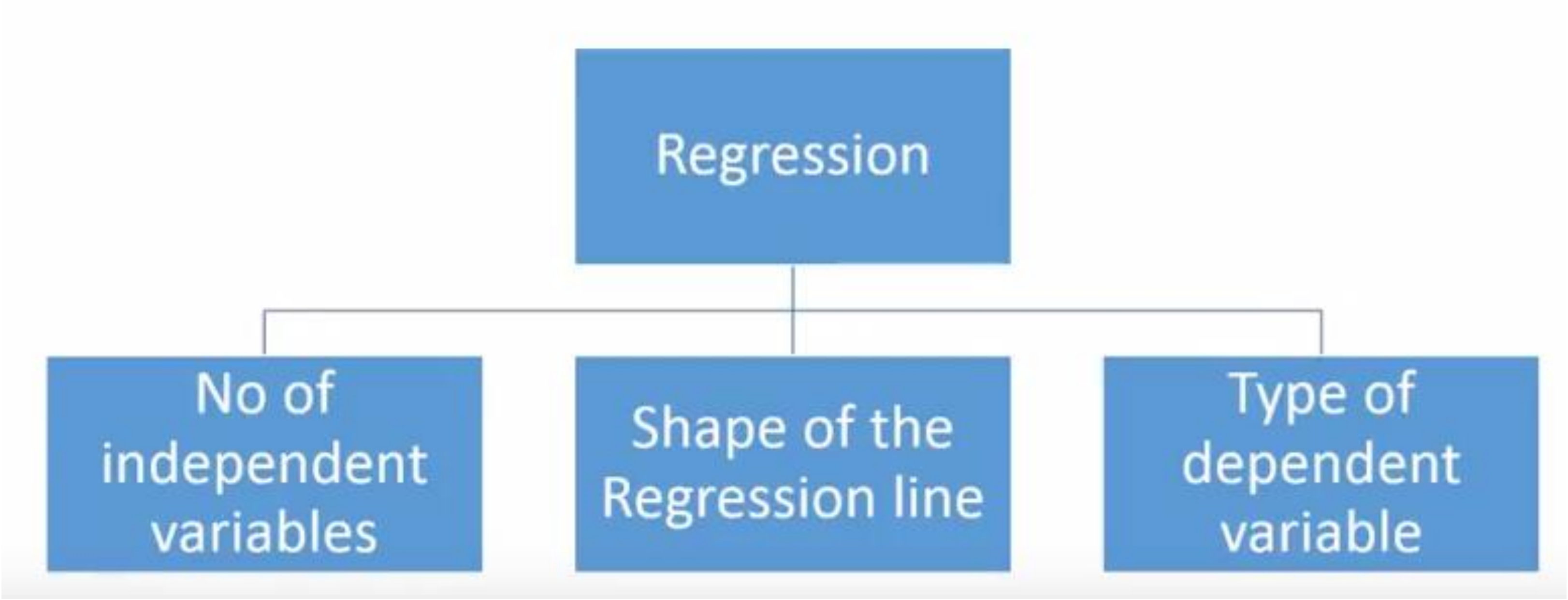
2

Predict the country from where the person comes from

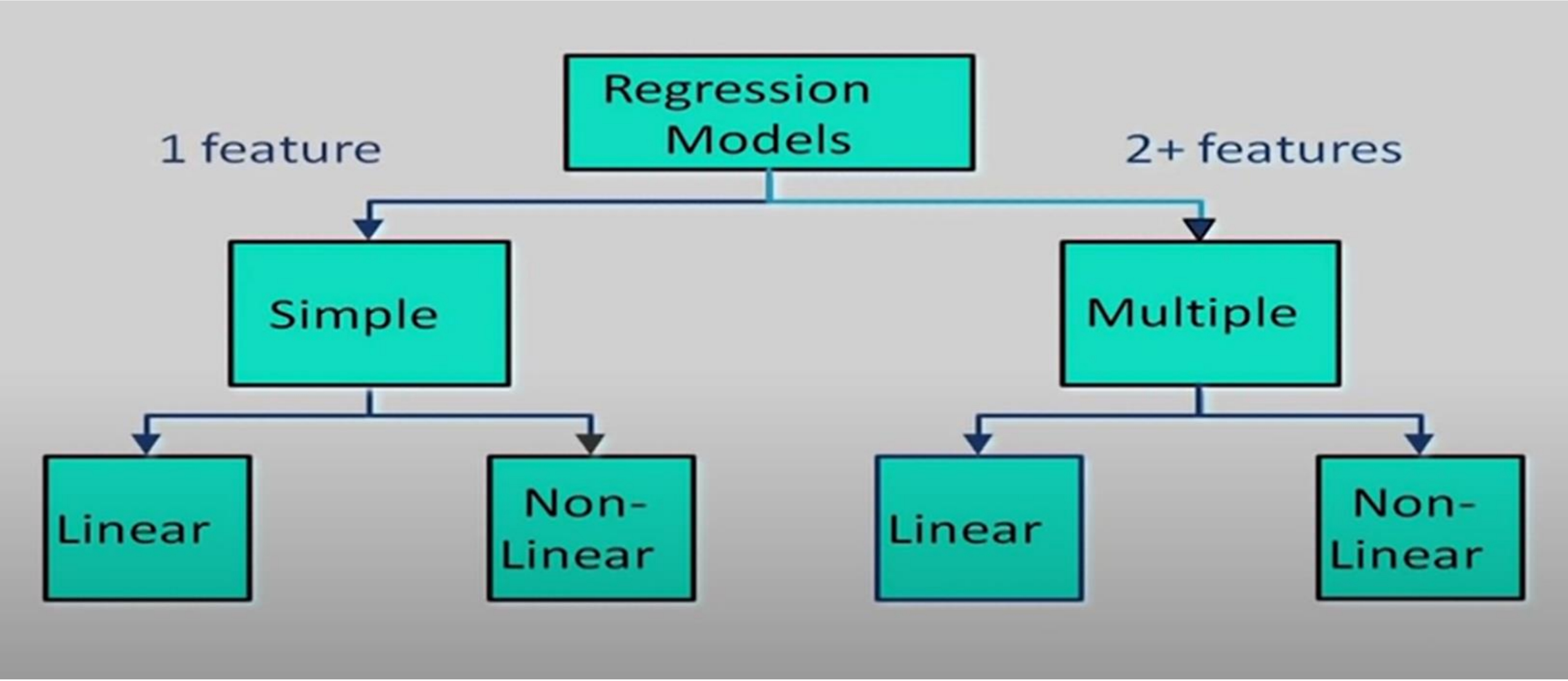
3

Predicting whether stock price of a company will increase tomorrow

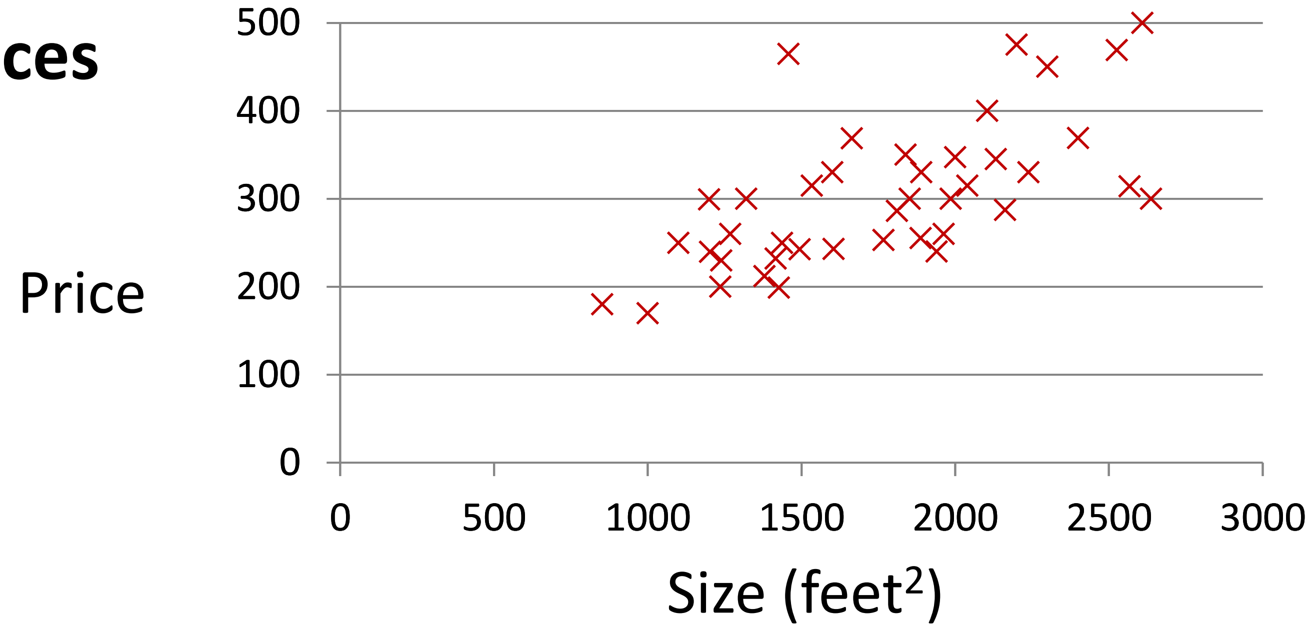
Types of regression models



Types of regression models



Housing Prices

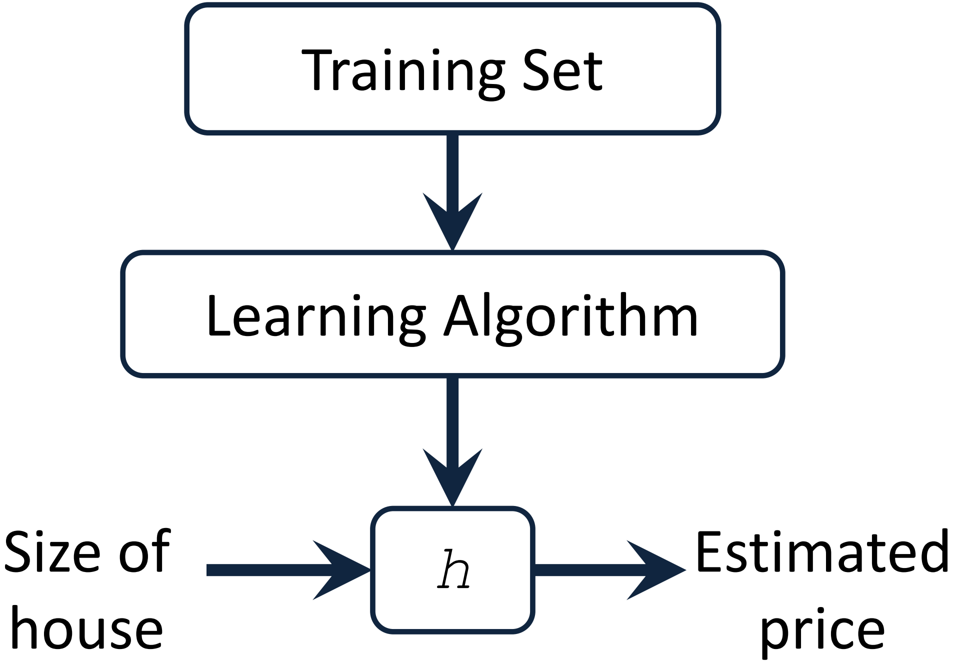


Supervised Learning

Given the "right answer" for each example in the data.

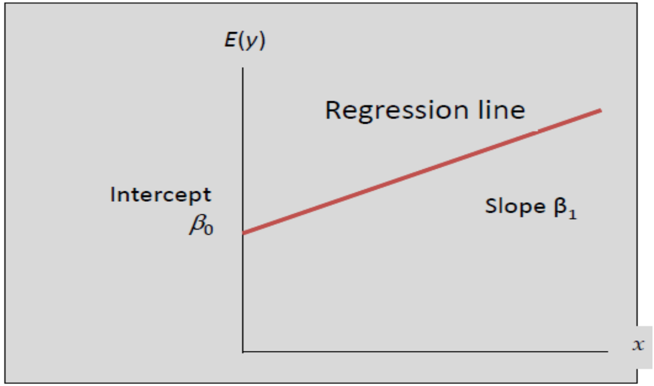
Regression Problem

Predict real-valued output



How do we represent h ?

$$h_\Theta(x) = \Theta_0 + \Theta_1 x$$



Linear regression with one variable.
Univariate linear regression.

Linear Regression Model

- Relationship Between Variables Is a Linear Function

Population Y-Intercept

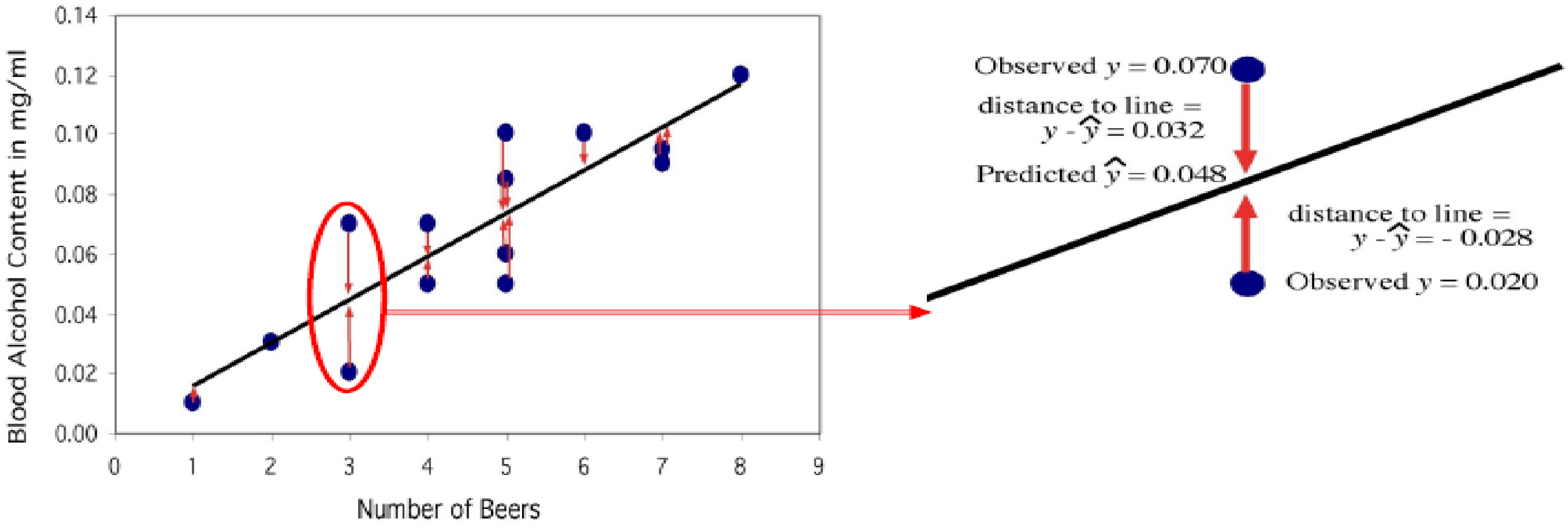
Population Slope

Random Error

$$Y = \beta_0 + \beta_1 x_1 + \epsilon$$

The regression line

The least-squares regression line is the unique line such that the sum of the squared vertical (y) distances between the data points and the line is the smallest possible.



Criterion for choosing what line to draw: method of least squares

- The method of least squares chooses the line ($\widehat{\beta_0}$ and $\widehat{\beta_1}$) that makes the sum of squares of the residuals $\sum \varepsilon_i{}^2$ as small as possible
- Minimizes

$$\sum_{i=1}^{n}[y_i - (b_0 + b_1 x_i)]^2$$

for the given observations (x_i, y_i)

How do we "learn" parameters

- For the 2-*d* problem

$$Y = \beta_0 + \beta_1 X$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

- To find the values for the coefficients which minimize the objective function we take the partial derivates of the objective function (SSE) with respect to the coefficients. Set these to 0, and solve.

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

$$\beta_0 = \frac{\sum y - \beta_1 \sum x}{n}$$

Example

Calculate the regression coefficient and obtain the lines of regression for the following data

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

X	Y	X^2	Y^2	X^Y
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
$\sum X = 28$	$\sum Y = 77$	$\sum X^2 = 140$	$\sum Y^2 = 875$	$\sum XY = 334$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{28}{7} = 4 ,$$

$$\overline{Y} = \frac{\Sigma Y}{N} = \frac{77}{7} = 11$$

Regression coefficient of X on Y

$$b_{xy} = \frac{N\Sigma XY-(\Sigma X)(\Sigma Y)}{N\,\Sigma Y^2-(\Sigma Y)^2}$$

$$= \frac{7(334)-(28)(77)}{7(875)-(77)^2}$$

$$= \frac{2338-2156}{6125-5929}$$

$$= \frac{182}{196}$$

$$b_{xy} = 0.929$$

(i) Regression equation of X on Y

$$X-\overline{X} = b_{xy}(Y-\overline{Y})$$

$$X-4 = 0.929(Y-11)$$

$$X-4 = 0.929Y-10.219$$

$\therefore$ The regression equation X on Y is $X = 0.929Y-6.219$

(ii) Regression coefficient *of* Y on X

$$b_{yx} = \frac{N\Sigma XY-(\Sigma X)(\Sigma Y)}{N\,\Sigma X^2-(\Sigma X)^2}$$

$$= \frac{7(334)-(28)(77)}{7(140)-(28)^2}$$

$$= \frac{2338-2156}{980-784}$$

$$= \frac{182}{196}$$

$$\therefore \quad b_{yx} = 0.929$$

(iii) Regression equation of Y on X

$$Y-\overline{Y} = b_{yx}(X-\overline{X})$$

$$Y-11 = 0.929\,(X-4)$$

$$Y = 0.929X-3.716+11$$

$$= 0.929X+7.284$$

The regression equation of Y on X is $Y = 0.929X + 7.284$

House Number	*Y*: Actual Selling Price	*X*: House Size (100s ft2)
1	89.5	20.0
2	79.9	14.8
3	83.1	20.5
4	56.9	12.5
5	66.6	18.0
6	82.5	14.3
7	126.3	27.5
8	79.3	16.5
9	119.9	24.3
10	87.6	20.2
11	112.6	22.0
12	120.8	.019
13	78.5	12.3
14	74.3	14.0
15	74.8	16.7

Important points about LR

- more susceptible to outliers hence;
- it should not be used in the case of big-size data.
- There should be a linear relationship between independent and dependent variables.
- There is only one independent and dependent variable.
- The type of regression line: a best fit straight line.

Advantages And Disadvantages of LR

Advantages	Disadvantages
Linear regression performs exceptionally well for linearly separable data	The assumption of linearity between dependent and independent variables

Advantages And Disadvantages of LR

Advantages	Disadvantages
Linear regression performs exceptionally well for linearly separable data	The assumption of linearity between dependent and independent variables
Easier to implement, interpret and efficient to train	It is often quite prone to noise and overfitting

Advantages and Disadvantages of LR

Advantages	Disadvantages
Linear regression performs exceptionally well for linearly separable data	The assumption of linearity between dependent and independent variables
Easier to implement, interpret and efficient to train	It is often quite prone to noise and overfitting
It handles overfitting pretty well using dimensionally reduction techniques, regularization, and cross-validation	Linear regression is quite sensitive to outliers Hence, it should not be used in the case of big-size data.

Multiple linear regression

- is used to estimate the relationship between **two or more independent variables** and **one dependent variable**
- **Example:**
- The selling **price** of a house can depend on the desirability of the **location**, the number of **bedrooms**, the number of **bathrooms**, the **year** the house was built, the **square footage** of the lot and a number of other factors
- The **height** of a child can depend on the **height of the mother**, the **height of the father**, **nutrition**, and environmental factors.

Multiple linear regression model

Multiple Regression Model with k Independent Variables:

Y-intercept

Population slopes

Random Error

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

Regression Model

- Our model assumes that

  $E(Y \mid X = x) = \beta_0 + \beta_1 x$ (the "population line")

Population line

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

Actual line from which the examples are drawn.

Least Squares line

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_p X_p$$

Estimated equation we want to create.

We use $\widehat{\beta_0}$ through $\widehat{\beta_p}$ as guesses for β_0 through β_p and $\widehat{Y}_i$ as a guess for Y_i. The guesses will not be perfect.

Multiple linear regression model

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + + \beta_k X_k + \xi$$

$$Y_i = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \xi$$

the n-tuples of observations are also assumed to follow the same model. Thus they satisfy

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} + \varepsilon_2$$
$$\vdots \qquad\qquad \vdots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + \varepsilon_n.$$

These n equations can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} \cdots x_{1k} \\ 1 & x_{21} & x_{22} \cdots x_{2k} \\ \vdots & \vdots & \vdots \quad \vdots \\ 1 & x_{n1} & x_{n2} \cdots x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or $y = X\beta + \varepsilon$.

- There is a closed form which requires matrix inversion, etc.
- There are iterative techniques to find weights
  - delta rule (also called LMS method) which will update towards the objective of minimizing the SSE.

Cost Function

$$J(b,w)= \frac{1}{2m}\sum_{i=1}^{m}\left(h_{b,w}(x^{(i)}) - y^{(i)}\right)^2$$

Find b and w to minimize J ($\min_{b,w} J(b,w)$)

Have some function $J(b,w)$

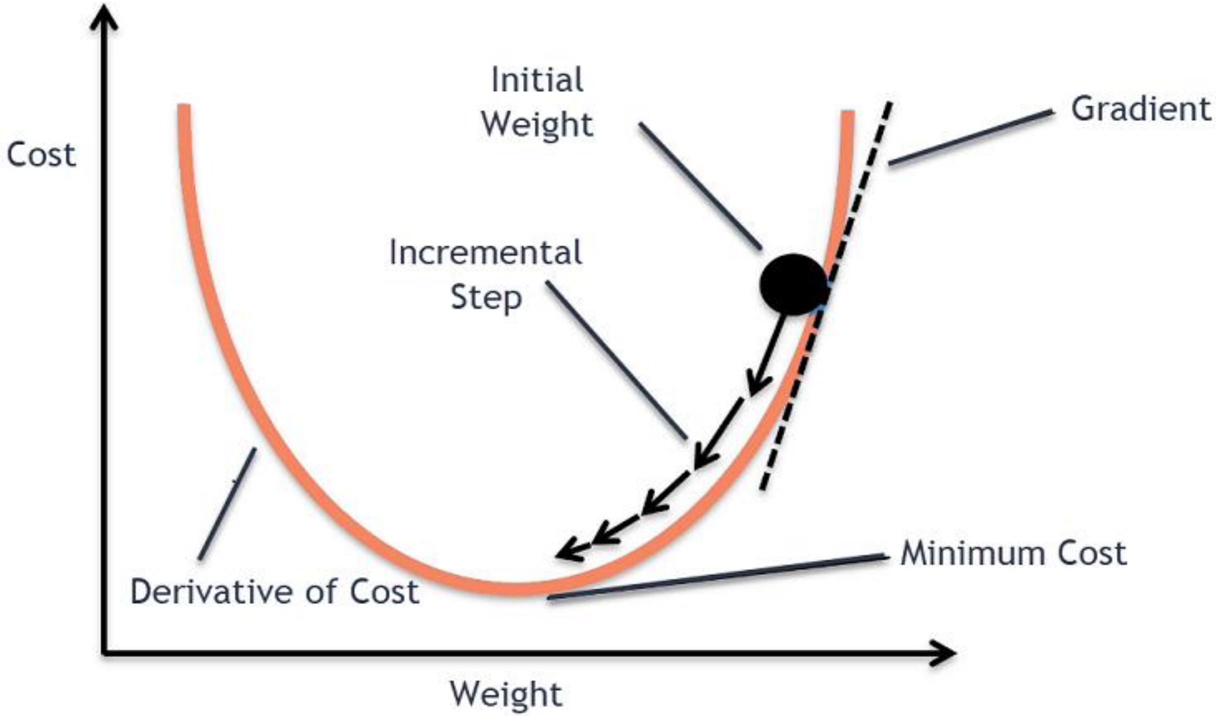
Want $\min_{b,w} J(b,w)$

Outline:

- Start with some b,w
- Keep changing b,w to reduce $J(b,w)$

  until we hopefully end up at a minimum

Gradient descent/delta rule/LMS method



Gradient descent algorithm

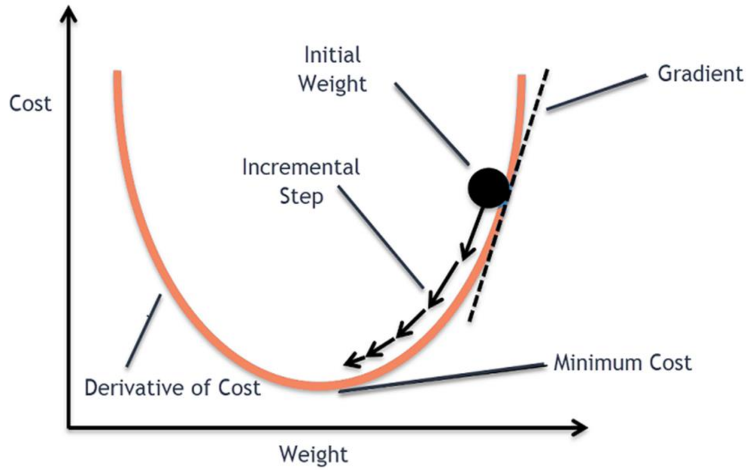
$$J = \frac{1}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i)^2$$

$$\frac{\partial J}{\partial a_0} = \frac{2}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i)$$

$$\frac{\partial J}{\partial a_1} = \frac{2}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i) \cdot x_i$$

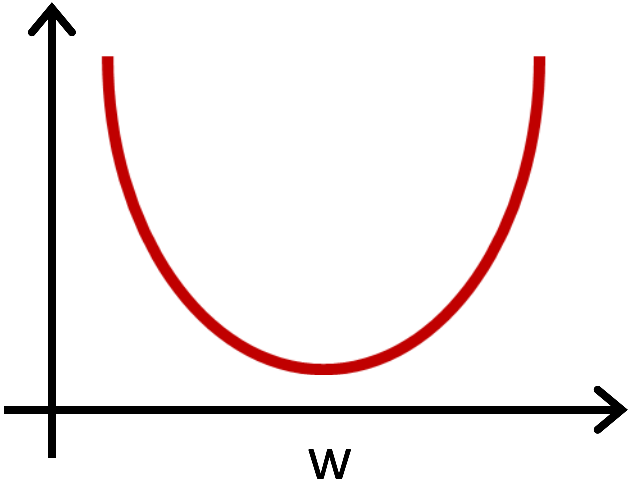
$$a_0 = a_0 - \alpha \cdot \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i)$$

$$a_1 = a_1 - \alpha \cdot \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i) \cdot x_i$$

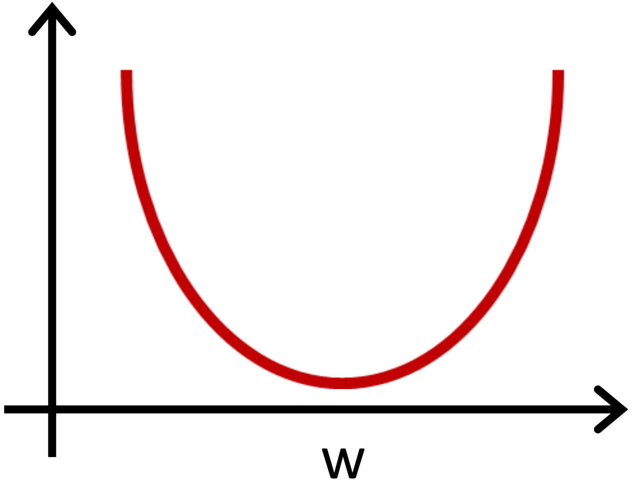


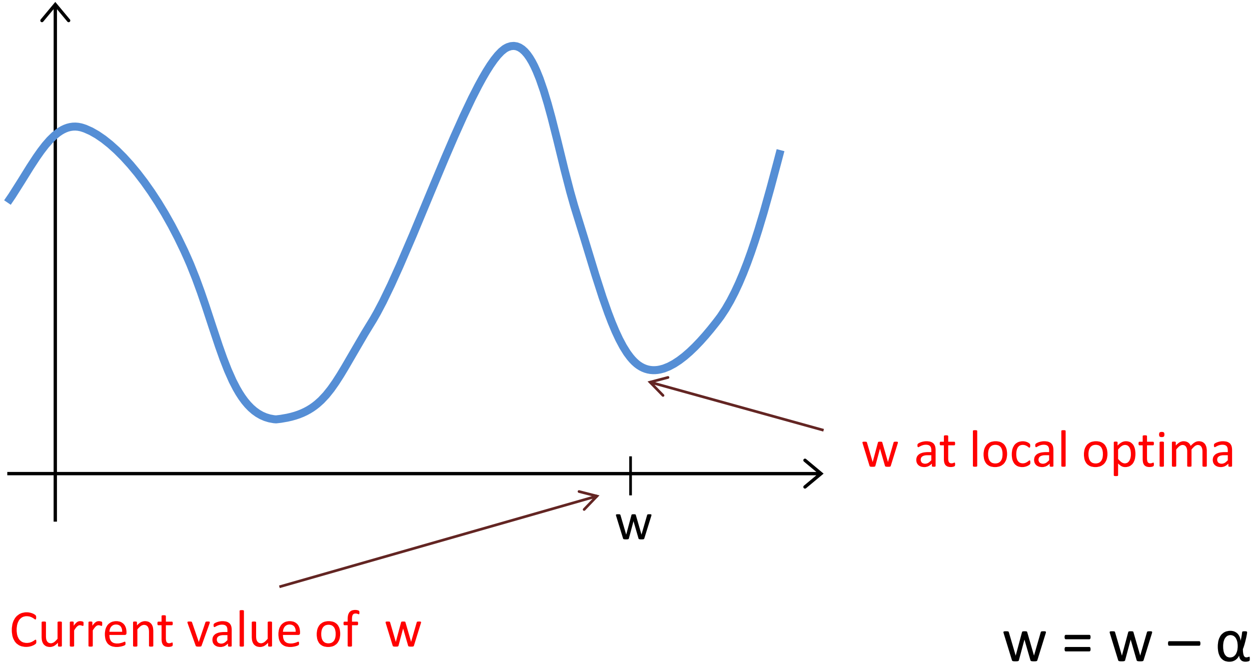
$$w = w - \alpha \frac{\partial}{\partial w} J(w)$$

If α is too small, gradient descent can be slow.



If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.



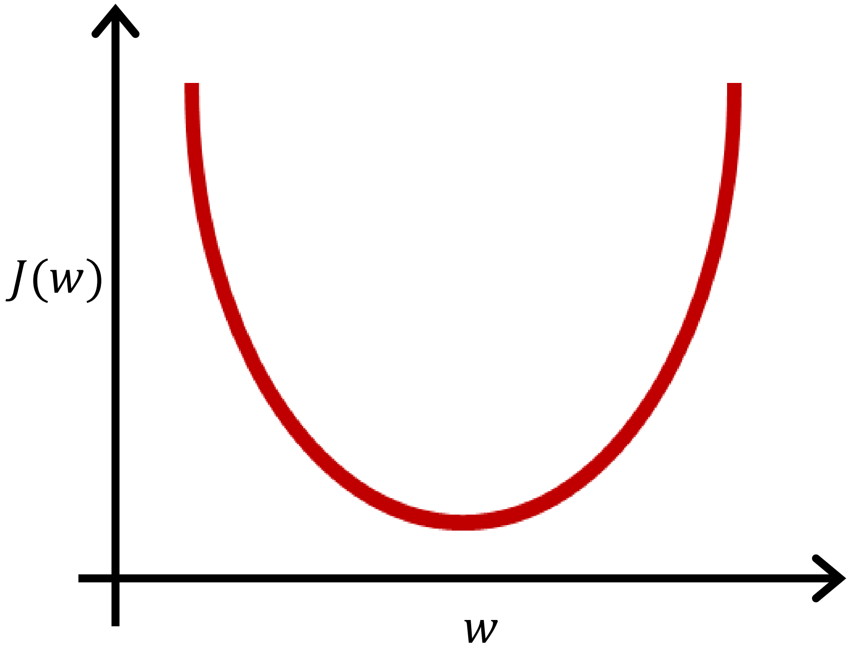


$$w = w - \alpha \frac{\partial}{\partial w} J(w)$$

Gradient descent can converge to a local minimum, even with the learning rate α fixed.

$$w = w - \alpha \frac{\partial}{\partial w} J(w)$$

As we approach a local minimum, gradient descent will automatically take smaller steps. So, no need to decrease α over time.



Benefits of Regression Analysis

- It provides a functional relationship between two or more related variables
- Helps in prediction and forecasting
- Improves business efficiency
- Supports business decisions
- Analyzing results and correcting errors
- Finding new opportunities

Limitations of Regression Analysis

- It is assumed that the cause and effect relationship between the variables remains unchanged. This assumption may not always hold good and may lead to misleading results.
- It involves very lengthy and complicated procedure of calculations and analysis
- It cannot be used in case of qualitative phenomenon viz. honesty, crime etc.