LINEAR REGRESSION FOR MACHINE LEARNING

PRESENTATION BY – SAURABH KOLHE

CONTENTS

- Basic Terminologies
- Regression-Definition
- Types of Regression
- Linear Regression
- Gradient Descent

DEPENDENT AND INDEPENDENT VARIABLES

Independent variables are regarded as input to system and may take on different values freely. It is also called as predictor, denoted by X.

Dependent variables are those which change as a consequence of change in other variables. It is also known as response variable, denoted by Y.

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39
144372.41	118671.85	383199.62	New York	182901.99
142107.34	91391.77	366168.42	Florida	166187.94
131876.9	99814.71	362861.36	New York	156991.12
134615.46	147198.87	127716.82	California	156122.51
130298.13	145530.06	323876.68	Florida	155752.6
120542.52	148718.95	311613.29	New York	152211.77
123334.88	108679.17	304981.62	California	149759.96
101913.08	110594.11	229160.95	Florida	146121.95
100671.96	91790.61	249744.55	California	144259.4
93863.75	127320.38	249839.44	Florida	141585.52
91992.39	135495.07	252664.93	California	134307.35
119943.24	156547.42	256512.92	Florida	132602.65
114523.61	122616.84	261776.23	New York	129917.04
78013.11	121597.55	264346.06	California	126992.93

DATASET SHOWING PROFIT TRENDS OF AN ORG.

Determine dependent and independent variables.

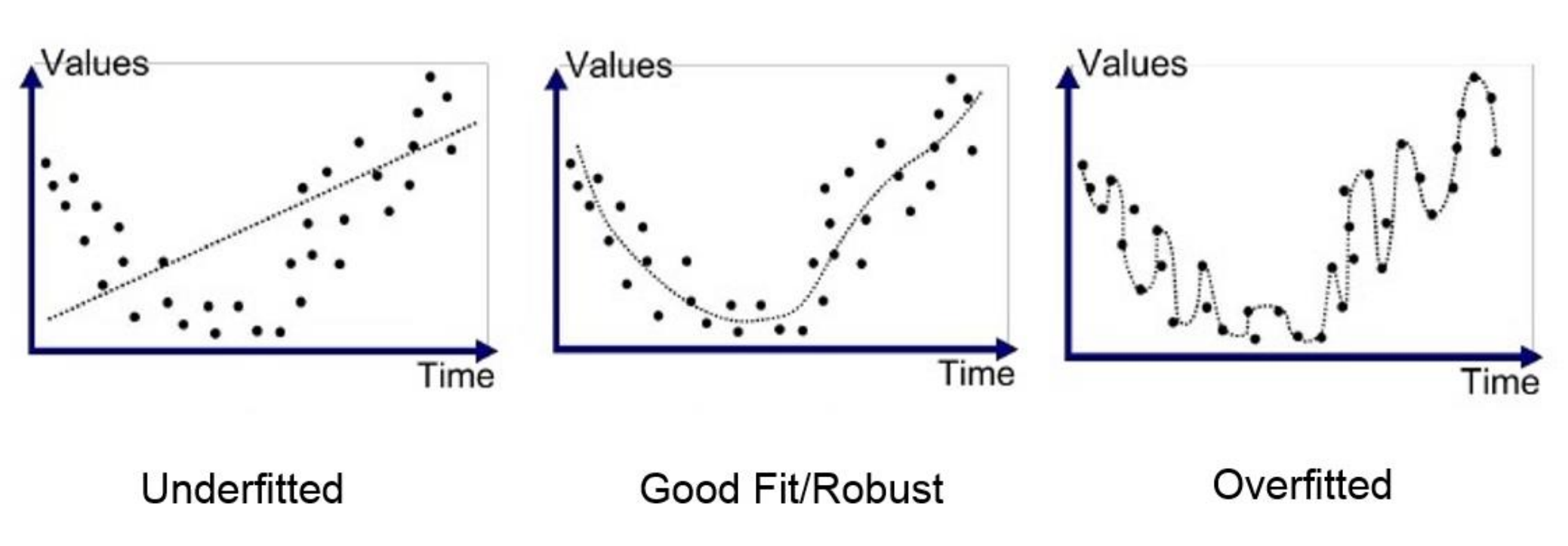
HYPERPARAMETER, EPOCH AND LEARNING RATE

In machine learning, a hyperparameter is a parameter whose value is set before the learning process begins.

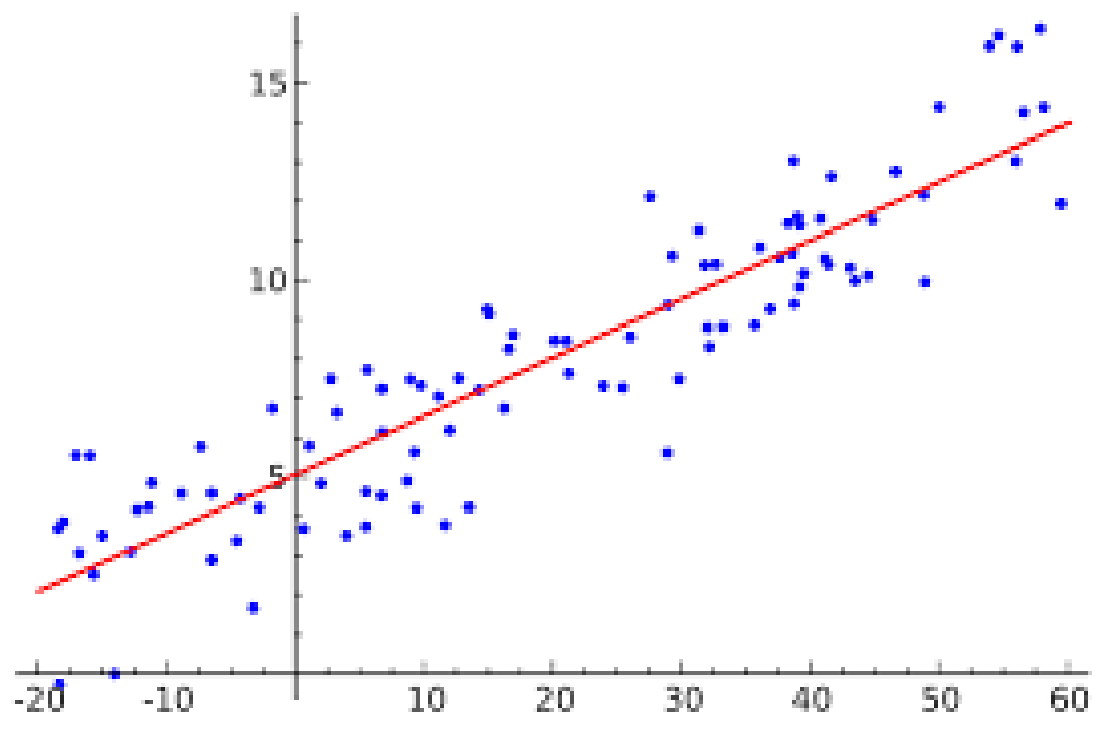
The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset.

The learning rate or step size in machine learning is a hyperparameter which determines to what extent newly acquired information overrides old information.

UNDERFITTING AND OVERFITTING



WHAT IS REGRESSION ?



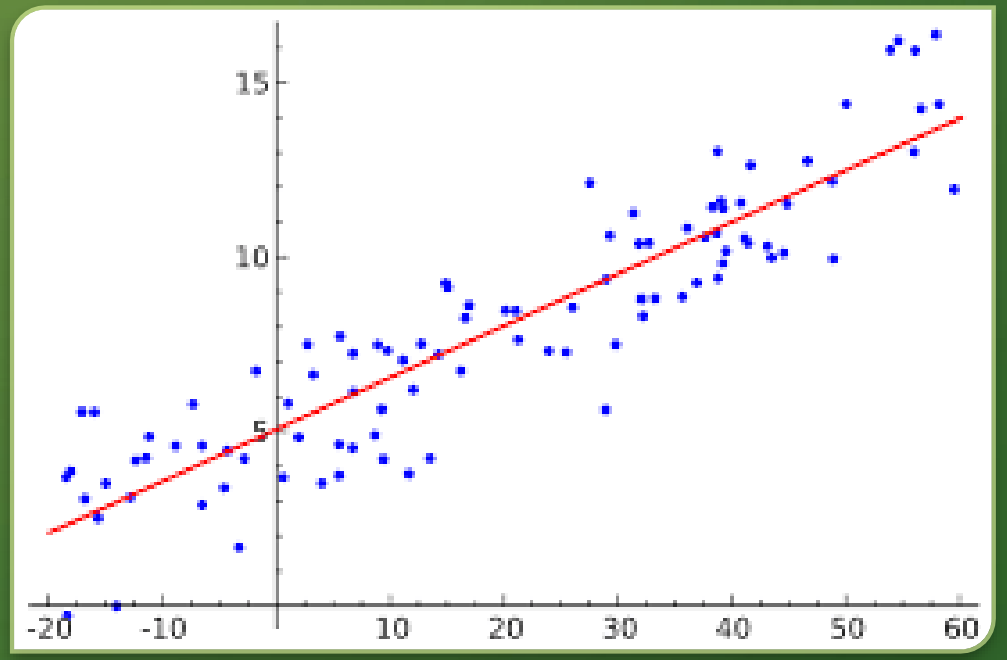
- It is a statistical measure that attempts to determine the strength of the relationship between one dependent variable(Y) and other changing independent variables(X).
- It is widely used for prediction and forecasting.

TYPES OF REGRESSION

- LINEAR REGRESSION – Univariate and Multivariate
- POLYNOMIAL REGRESSION – Univariate and Multivariate
- LOGISTIC REGRESSION – Univariate and Multivariate

LINEAR REGRESSION

- In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).



YearsExperience	Salary
1.1	39343
1.3	46205
1.5	37731
2	43525
2.2	39891
2.9	56642
3	60150
3.2	54445
3.2	64445
3.7	57189
3.9	63218
4	55794

UNIVARIATE LINEAR REGRESSION

A model consisting of one independent and one dependent variable.

Equation:

$y = mx + c$

R&D Spend	Administration	Marketing Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83
162597.7	151377.59	443898.53	California	191792.06
153441.51	101145.55	407934.54	Florida	191050.39
144372.41	118671.85	383199.62	New York	182901.99
142107.34	91391.77	366168.42	Florida	166187.94
131876.9	99814.71	362861.36	New York	156991.12
134615.46	147198.87	127716.82	California	156122.51
130298.13	145530.06	323876.68	Florida	155752.6
120542.52	148718.95	311613.29	New York	152211.77
123334.88	108679.17	304981.62	California	149759.96
101913.08	110594.11	229160.95	Florida	146121.95
100671.96	91790.61	249744.55	California	144259.4
93863.75	127320.38	249839.44	Florida	141585.52
91992.39	135495.07	252664.93	California	134307.35
119943.24	156547.42	256512.92	Florida	132602.65
114523.61	122616.84	261776.23	New York	129917.04
78013.11	121597.55	264346.06	California	126992.93

MULTIVARIATE LINEAR REGRESSION

A model consisting of more than one independent variables.

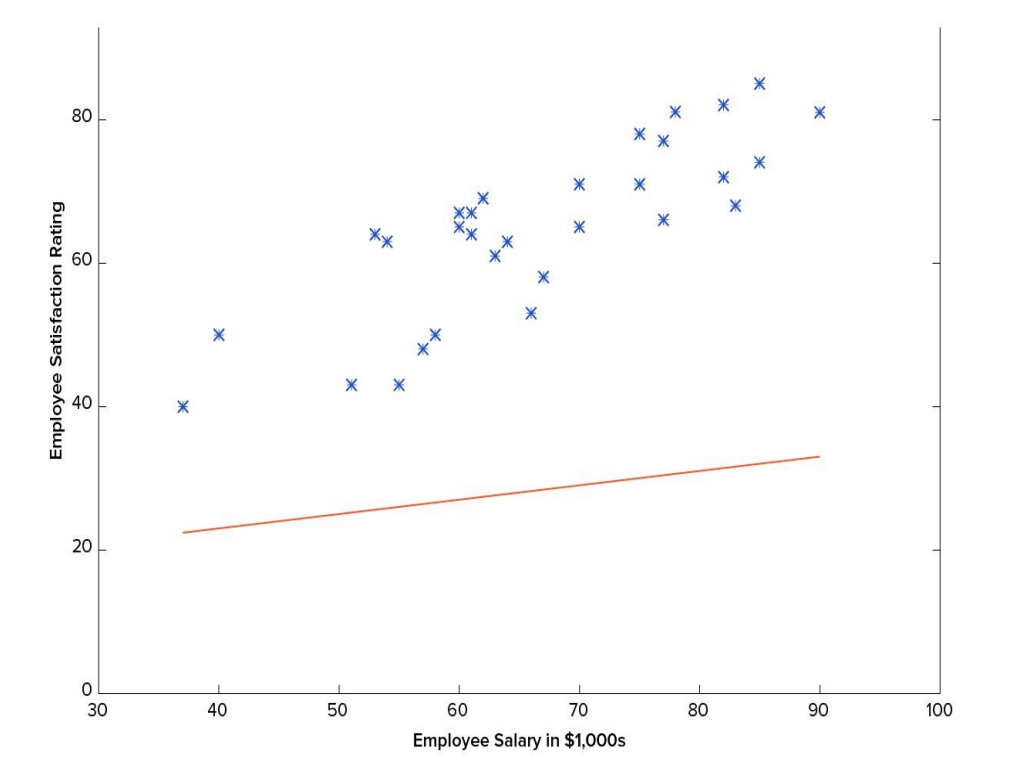
Equation:

$$y = m_1x_1 + m_2x_2 + m_3x_3.... + c$$

REAL WORLD EXAMPLES

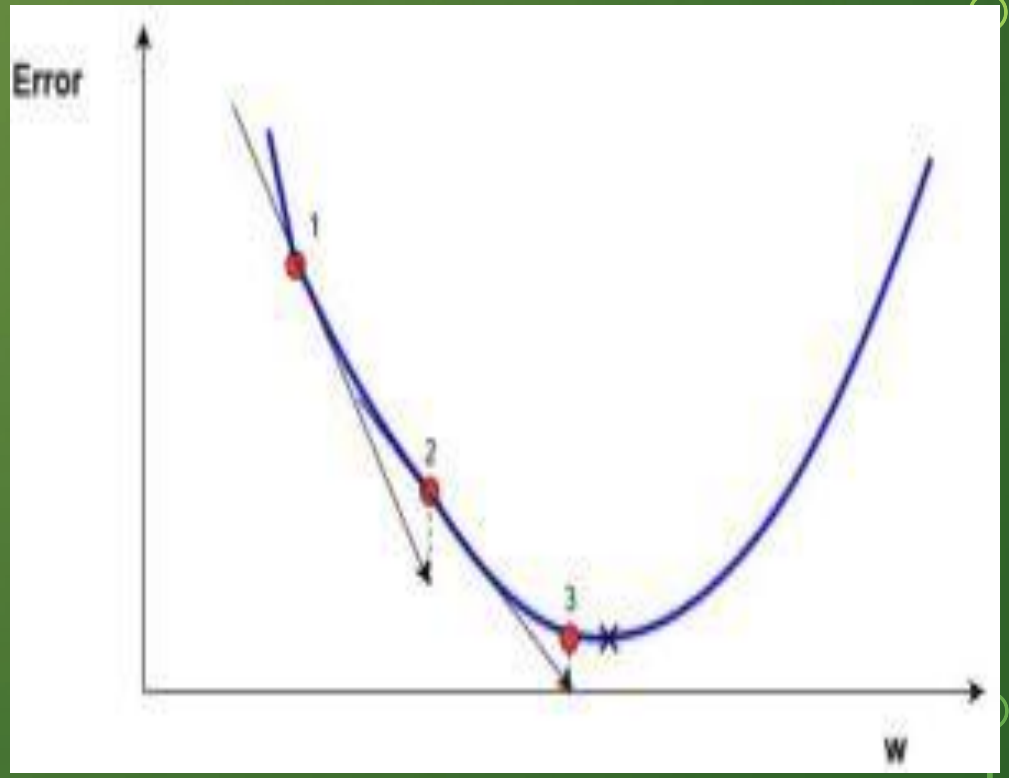
- Linear Regression is a rather ubiquitous curve fitting and machine learning technique that's used everywhere from scientific research teams to stock markets.
- Studying engine performance from test data in automobiles.
- Least squares regression is used to model causal relationships between parameters in biological systems.
- OLS regression can be used in weather data analysis.
- Linear regression can be used in market research studies and customer survey results analysis.

HOW TO FIND VALUES OF PARAMETERS ?



GRADIENT DESCENT

- MSE : $y = (1/n)\sum_{i=1}^{n}(y_i - y_{pi})^2$
- Partial Differentiation w.r.t every parameter:
- d(m) = $(-2/n)\sum_{i=1}^{n}(y_i - y_{pi})x_i$
- d(c) = $(-2/n)\sum_{i=1}^{n}(y_i - y_{pi})$
- Update values of every parameter:
- m = m – L*d(m)
- c = c – L*d(c)
- Repeat epoch number of times.



R2 SCORE

- R-squared is a statistical measure of how close the data are to the fitted regression line.
- $R^2 = 1$ - SSr/SSt
- $R^2 = 1 - (\sum_{i=1}^{n}(yi - ypi)^2)/(\sum_{i=1}^{n}(yi - y_{avg})^2)$

THANK YOU