UNIT 3 - REGRESSION

Introduction

↳ Regression is Supervised Learning algo (ie learns from labelled data) to make predictions.

↳ focuses on predicting continuous outcomes based on relationship between independent & dependent (target) variables.

Independent → influence the dependent variable (size of house)

dependent → outcome that model aims to predict (house prize, temp.)

$y = f(x)$

(GRE Score) Dependent ← y ; x → independent (CGPA)

Types —

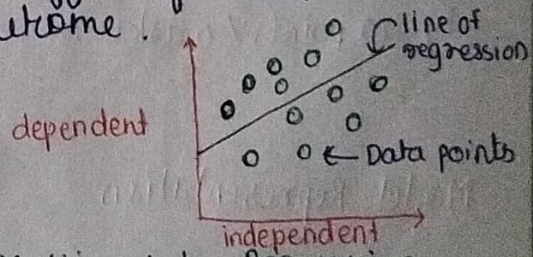
Linear Regression - models the relationship in straight line.

simple Linear Regression - has only 1 independent variable & 1 dependent

Multiple Linear Regression - has multiple independent var & 1 dependent var

SLR	MLR
① 1 independent var & 1 depend var.	① many independent var & 1 dependent var.
② $Y = \beta_0 + \beta_1 X + \epsilon$	② $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$
③ 1 independent var	③ 2 or more independent var.
④ Less complex	④ More complex.
⑤ Analyse effect of single factor on outcome	⑤ Analyse effect of multiple factors on outcome.

Polynomial Regression
Logistic Regression
Decision Tree Regression, etc.



Linear Regression cont as Univariate & Multivariate Regression + Bivariate

Univariate	multivariate
① Examines single variable at a time (dependent / independent).	① Examines multiple variable at a time.
② Less complex.	② More complex.

Similar to simple & multiple regression

Univariate Regression

↳ goal is to find relationship betⁿ single feature & target variable
↳ most common technique used is Simple Linear Regression, that fits a straight line to the data.

Least Square Method.

↳ used to find the best-fitting line for set of data points.
↳ minimizes the difference betⁿ actual & predicted values/data.

Steps

① We have pairs of data points (x, y) — x: independent, y: dependent

② Calculate difference betⁿ actual Y and predicted Y

③ Sum of all squares = $\Sigma (y_i - \bar{y}_i)^2$ — y_i: actual, $\bar{y}_i$: predicted

④ Find the best fit line by slope & intercept calculation

⑤ Final equation in form $Y = mx + c$ — x: independent, Y: dependent, m: slope, c: y-intercept

---

eg X - hours studied
Y - exam scores

∴ for every additional hour studied score increases by 10 points / marks.

① calculate sum/mean of X & Y, XY & X^2
② find slope & intercept
③ $Y = mx + c$ or $Y = \beta_1 x + \beta_0$
eg $Y = 50x + 10x$

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

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

Model Representation

Cost functions / Loss function

↳ calculates difference betⁿ predicted values & the actual values.

↳ minimizing cost function improves accuracy. ↓cost fn ↑accuracy

MSE (Mean Squared Error)

↳ average of squared differences betⁿ actual & predicted values.

$$MSE = \frac{1}{n} \Sigma (y\text{-actual} - y\text{-predicted})^2$$

↓ MSE = better fit

Applⁿ - evaluate model performace & make comparisons betⁿ different models

MAE (Mean Absolute Error)

↳ average magnitude of errors

↓ MAE = better fit

$$MAE = \frac{1}{n} \Sigma \, |y\text{-actual} - y\text{-predicted}|$$

MAE treat all erros equall not sensitive to outliers

Applⁿ - model accuracy, esp when outliers are present.

R^2 / Coefficient of Determination

↳ proportion of variance in dependent variable which can be explained by independent variable (ie can be predicted)

$$R^2 = 1 - \frac{\Sigma (y\text{-actual} - y\text{-predicted})^2}{\Sigma (y\text{-actual} - y\text{-mean})^2}$$

y-mean → mean of actual values

Ranges from 0 to 1

↑ value = better fit

SSE (Sum of Squared Error)

$$SSE = \Sigma (y\text{-actual} - y\text{-predicted})^2$$

Optimizing SLR with Gradient Descent Algorithm

↳ iterative optimization algorithm used to minimize the cost/loss function

↳ updates parameters (β_0 & β_1) in direction that reduce loss

Steps –

① Initialize $\beta_0 = \beta_1 = 0$

② Define learning rate 'L' to a small value like 0.01

L → determines how big of a step we take during each iteration.

③ Calculate gradient

$$D_{\beta_0} = \frac{-2}{n} \Sigma (y\text{-actual} - (\beta_1 x_i + \beta_0)) = \frac{-2}{n} \Sigma (y\text{-actual} - y\text{-pred})$$

~~$\frac{2}{n} \Sigma$ (y-actual − y-pred)~~

$$D_{\beta_1} = \frac{-2}{n} \Sigma (y\text{-actual} - (\beta_1 x_i + \beta_0)) x_i$$

$$= \frac{-2}{n} \Sigma x_i (y\text{-actual} - y\text{-pred})$$

④ Update β_0 & β_1 using gradients

update $\beta_0 = \beta_0 - L D_{\beta_0}$

update $\beta_1 = \beta_1 - L D_{\beta_1}$

⑤ Repeat 3 & 4 until cost function is small (ideally 0)

Multivariated Regression

↳ models the relationship between multiple independent & ~~multiple~~ multiple / one depedent variables

dependent $Y = \beta X + \epsilon$ → matrix of error terms
↳ matrix coefficient that represent relation bet^n predictors & outcomes
↳ independent

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & & & \\ \vdots & & & \\ y_{n1} & & \cdots & y_{np} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & & & & \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} \begin{bmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \beta_{11} & & \\ \vdots & & \\ \beta_{q1} & \cdots & \beta_{qp} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & & \\ \vdots & & \\ & & \epsilon_{np} \end{bmatrix}$$

Advantages -

Captures complex Relationship - models how multiple predictors affect 1 outcome

Improve predictive power
Identify important factors
Control for other variable
flexible

Disadvantages — complexity
risk of overfitting
Data Requirements
Assumes linearity

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n x_n + \epsilon$$

Polynomial Regression

↳ models relationship between independent & dependent variable as 'nth'- degree polynomial.

↳ it is extension of linear regregression by introducing polynomial terms, allows model to fit non-linear relationships.

Model Representation

$$Y = \beta_0 + \beta_1 X + \beta_2 X^{02} + \beta_3 X^3 + \cdots \beta_n X^n + \epsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & & & \\ 1 & x_3 & & & \\ \vdots & & & & \\ 1 & x_n & & & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Generalization

Bias - prediction ~~model~~ error that was introduced in model due to oversimplifying ML algorithm.
gap bet^n predicted & actual - represents difference between predicted and actual value

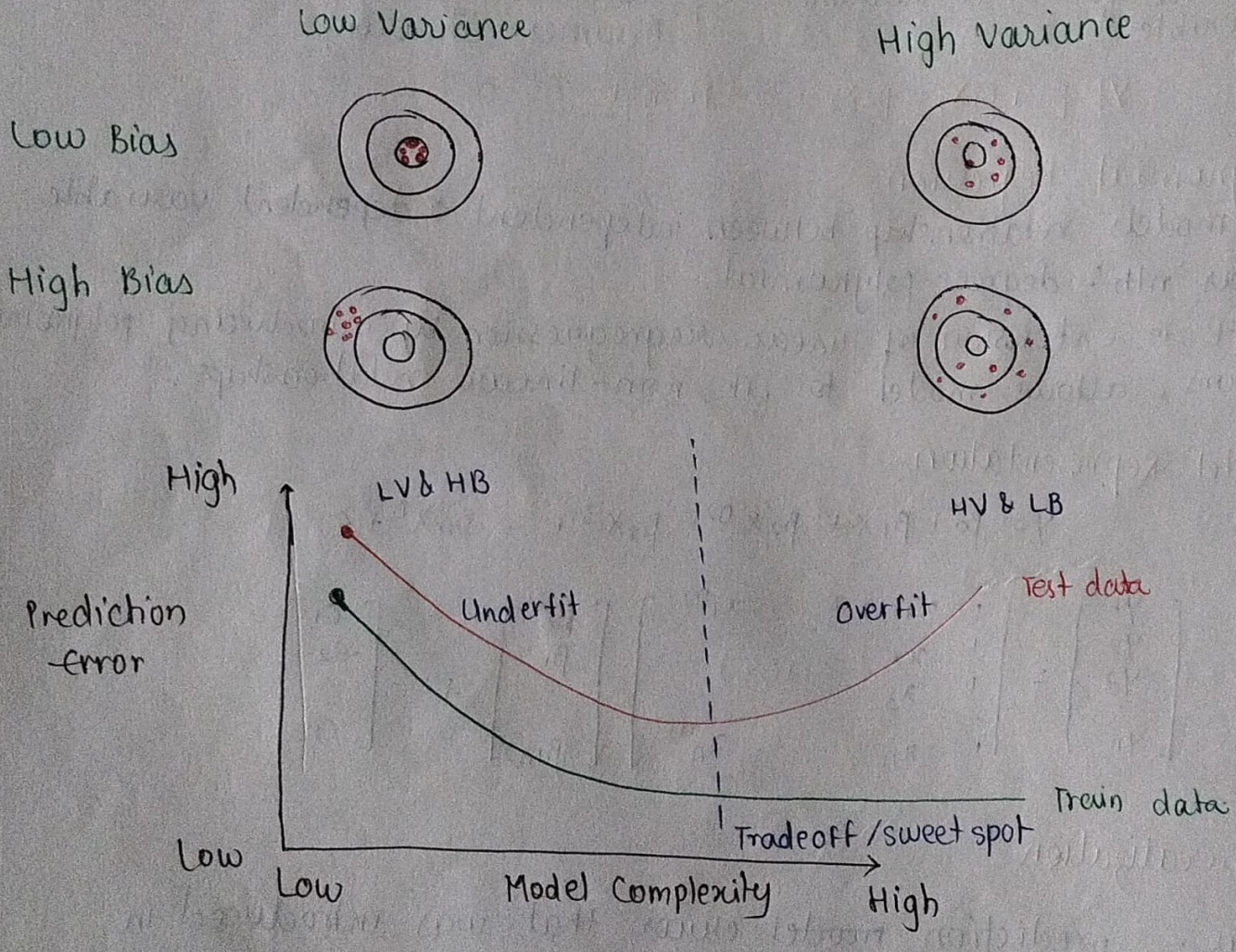
- High Bias = model is too simple and underfits the data

Variance - if model perform well with training dataset, but doesnot perform well with testing dataset, variance occurs
- it refers to models sensitivity to fluctuations in training data.
- high variance = model is overfitting training data & not perform well on test/unseen data.

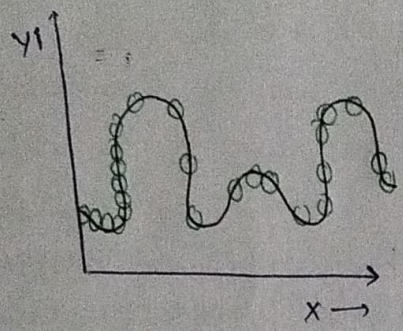
how much scattered predicted values are wrt each other

Bias - Variance Dilemma / Tradeoff

↳ trade-off betn 2 types of errors that affects the performance of predictive models.

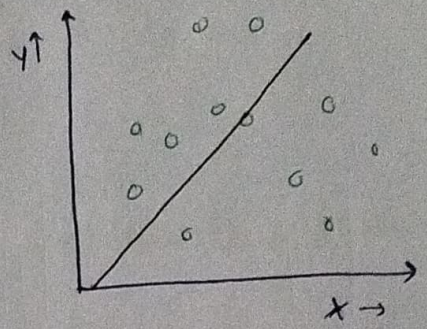


Overfitting & Underfitting



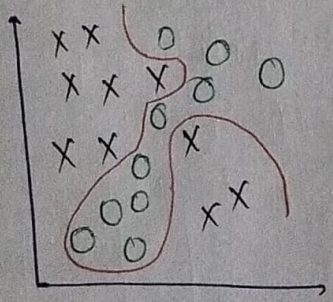
error occurs when model is too complex & captures noise in training data

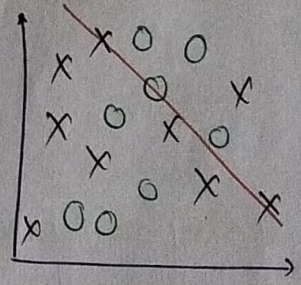
low training error
High testing error
Low bias
High variance
Too complex model complexity
Excellent performance on training data



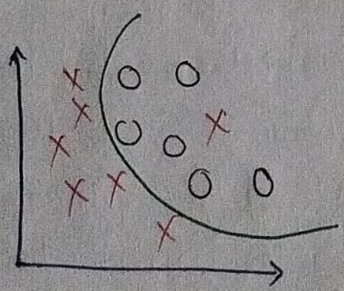
error occurs when model is too simple to capture the underlying trend of data

High training error
High testing error
High Bias
Low variance
Too simple model complexity
Low/Poor performance on training data





Appropriate fit - not underfit, niether overfit



Ways to prevent overfitting
↳ early stopping
↳ train with more data
↳ feature selection
↳ cross validation
↳ data augmentation
↳ Regularization.