25/3 unit 3 ML

(1) Training error & Generalization Error

↓

→ error calculated in training phase
→ error in known data set.

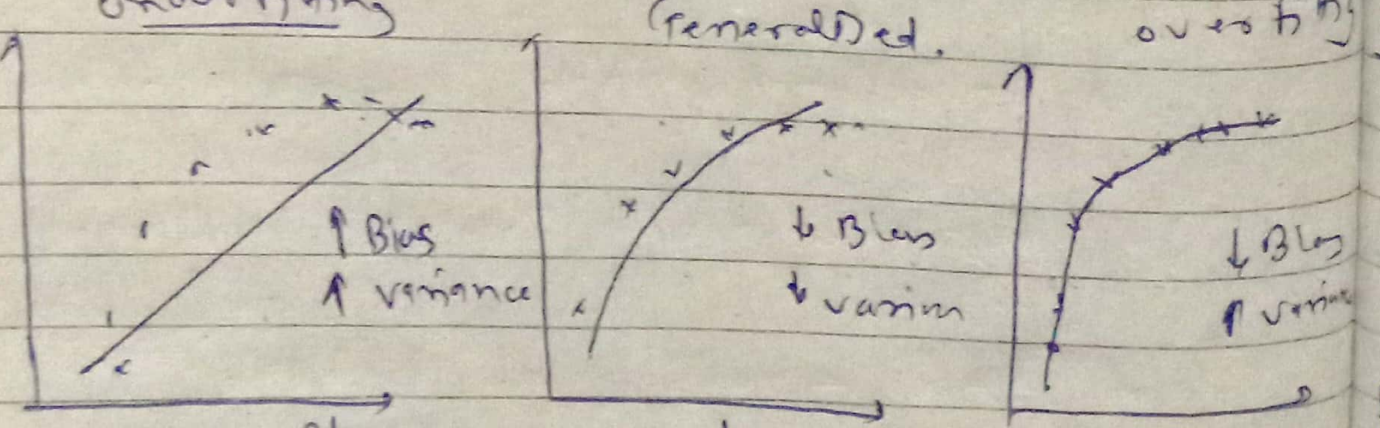
$$\varepsilon = \frac{1}{n} \sum_{1}^{n} \text{error}(f_D(x_i), y_i)$$

→ error calculated in testing phase
→ error in unknown or unseen dataset

<u>Bias</u> – algorithm error rate on the training set

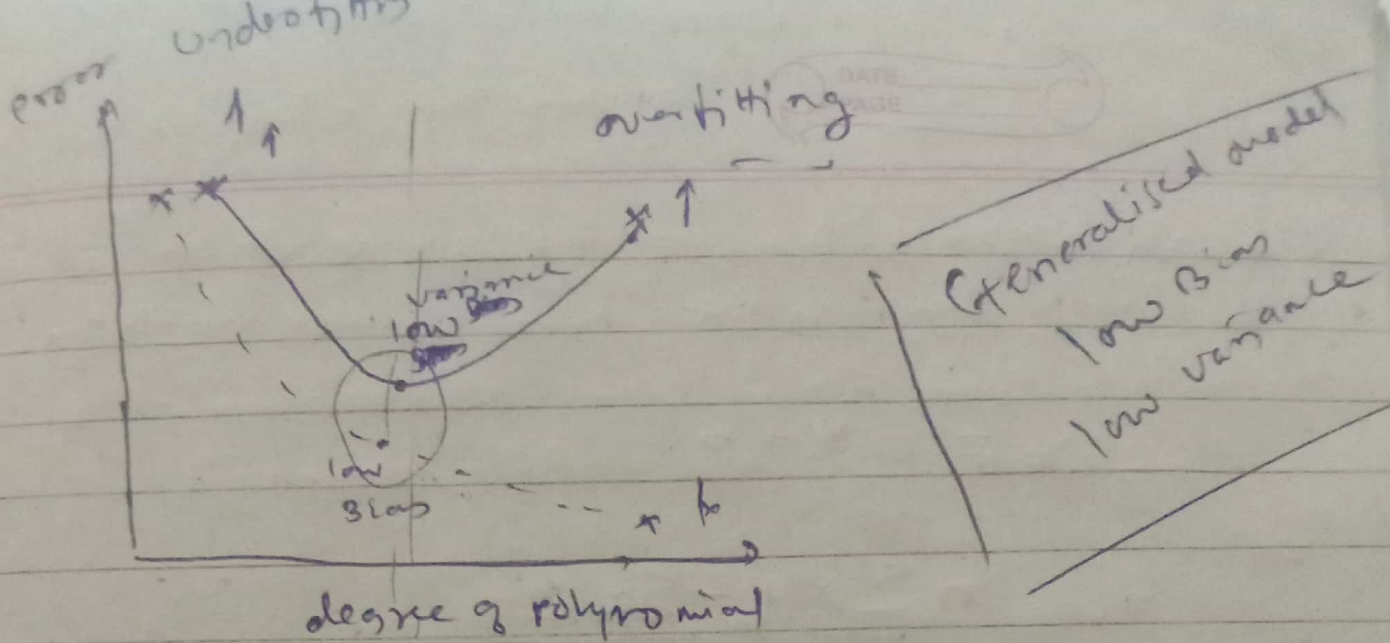
variance – After getting low error in training, we apply for testing and got high error, this is variance.
→ test set error.

underfitting | Generalised. | overfitting



polynomial = 1 degree | 2 degree | 4 degree

~~Linear~~ Regression



Underfitting
(1) Can't capture underlying Pattern
(2) does not fit well
(3) less data to build an accurate model
(4) ↑ Bias, ↑ variance

Reduce by (1) increase model complexity
(2) increase no. of feature
(3) Remove noise
(4) increase the no. of epochs.

Overfitting
(1) when we train we lot of data, and
Start learning noise & inaccurate data
(2) low training error But high testing error
(3) low Bias & high variance.

Reduce By (1) increase training data
(2) Reduce Complexity
(3) Regularization.

Simple & Multiple Regression
one variable
More than 2 variable

Linear Regression — Linear function
Non Linear Regression — Non Linear function

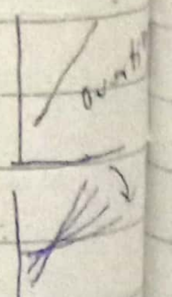
$$y' = ax + b$$

$$a = \frac{n\sum xy - \sum x \sum y}{n \cdot \sum x^2 - (\sum x)^2}$$

$$b = \frac{1}{n}\left(\sum y - a \times \sum x\right) = $$

LASSO Regression (L1 Regularization)

→ Regularization technique
→ overfitting issues solved
→ feature selection.
→ used in multicollinearity.



Regularization → adding penalty to the best fit

penalty = absolute value of Magnitude of the coeffi...

cost function ⇒
Residual sum of Square + λ * (sum of the absolute value of the Magnitude of Coefficient)

we try to minimize this

$$\sum_{i=1}^{n} \left[Y_i - \sum x_{ij} \beta_j \right]^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

λ = amount of shrinkage (Cross validation is used to find λ)

Ridge Regression (L2 regularization)

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

penalizing

$$\left(h(x)^{(i)} - y^{i} \right)^2 + \lambda\,(\text{slope})^2$$

$$= 0 + 1(2)^2$$

$$= 4 \downarrow\downarrow\downarrow$$

① Overfitting Reduce

$$\sum_{i=1}^{n} \left[Y_i - \sum x_{ij} \beta_j \right]^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

② prevent multicollinearity.

Gradient Descent Algorithm

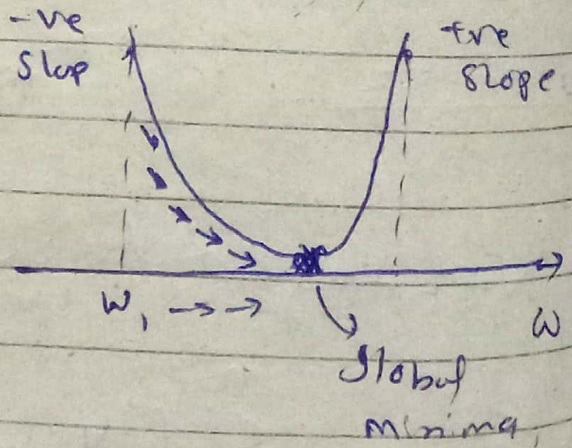
→ optimization algorithm - Minimizing the cost function by the Model's parameter.

→ using iteration.
→ It helps in finding the local minimum of a function.

$$W_{new} = W_{old} - \eta \times \frac{\partial L}{\partial W_{old}}$$ → n data point Gradient descent =

$$\text{loss} = \sum_{i=1}^{x} (y - \hat{y})^2$$

$$= \sum_{i=1}^{n} (y - \hat{y})^2$$



n data point - GD
1 data point SGD
k data point - Minibatch GD.

Evaluation Metrics. (Cost function).

to Build and deploy a generalized Model we require to evaluate the Model on different Metrics which helps us

* Mean error → average of the diff. b/w the actual value & predicted value

* Mean Squared error - (MSE)
→ Squared average of Square of diff. b/w the actual value & predicted value

Mean Absolute error (MAE)
$$\rightarrow \frac{1}{N} \sum |y - \bar{y}|$$

Root Mean Squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{N} \sum (y - \bar{y})^2}$$

R^2 (R Squared)

$$R2 = \frac{SSr}{SSm}$$

SSr = Squared Sum error of regression line

SSm = Squared sum error of Mean line

DATE
PAGE

R^2 Show how well the data fit the regression Model.