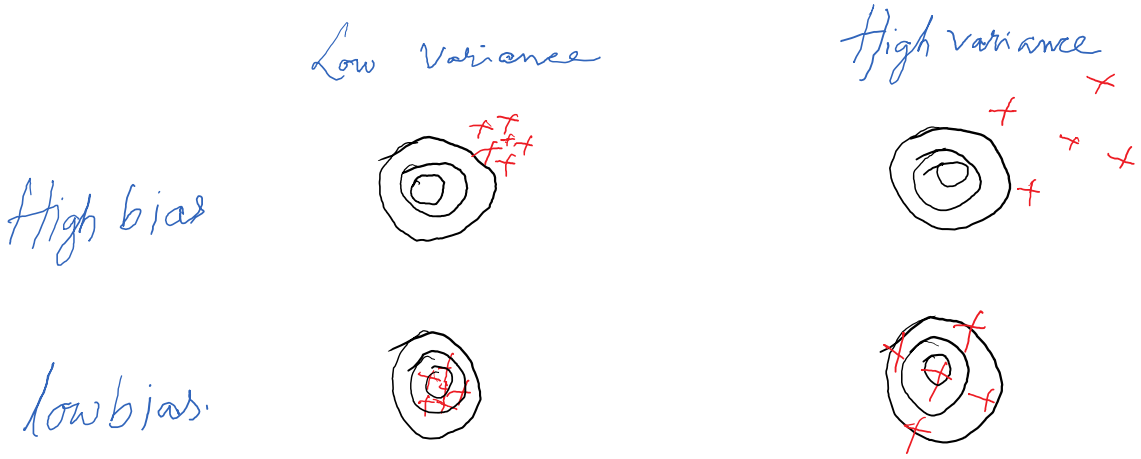
Cross-Validation & Bias-Variance Tradeoff & Linear Model Selection



$$Error(x) = Noise(x) + bias(x) + Variance(x)$$

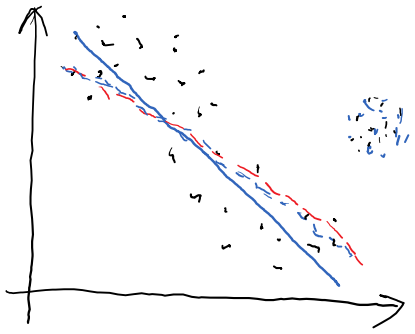
Bias → Algorithm tendancy to continuously learn the wrong way by not taking in to account all the info in the data.

(UNDER FIT)

Variance → Algorithm tendancy to learn Random things irrespective of the real signal, by fitting a highly flexible model that is following error/Noise very closely (OVER FIT)

→ Bias → How much average accuracy of model changes as change in i/p data.

→ Variance → How sensitive the algo is to change in data!

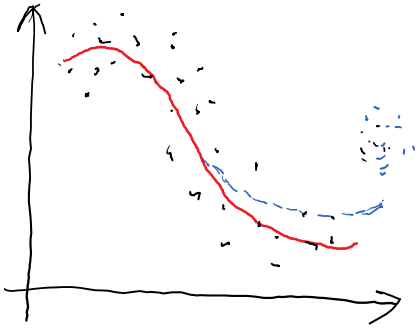


$y = mx + c$

Degree = 1

HB
LV

Under fit

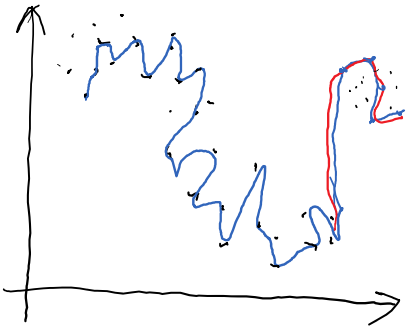


$y = m_1 x_1 + C + m_2 C_2^2 + m_3 C_3^3$

Degree = 3

Ideal fit

polynomial eqn



$y = C + m_1 x_1 + m_2 x_2^2 + m_3 x_3^3 + \ldots m_{20} x_{20}^{20}$

Degree = 20

Over-fit

LB
HV

Linear model selection & Regularisation

→ Alternative fitting procedure instead of "Least square"

→ Why → Improve the prediction Accuracy

→ Model Interpretability

Linear = $y = \beta_0 + \beta_1 x_1' + \beta_2 x_2' + \beta_3 x_3' + \ldots + \epsilon$

① Prediction accuracy — n → data, p → variables

→ Least square → low bias & low variance $[n \gg p]$

→ Variance ↑ → Overfit $[n \not\gg p]$

→ ∞ Variance → Not useful $[p > n]$

By Shrinking / Constraining the Co-efficients

→ Highly reduce the variancy with negligible increase in bias.

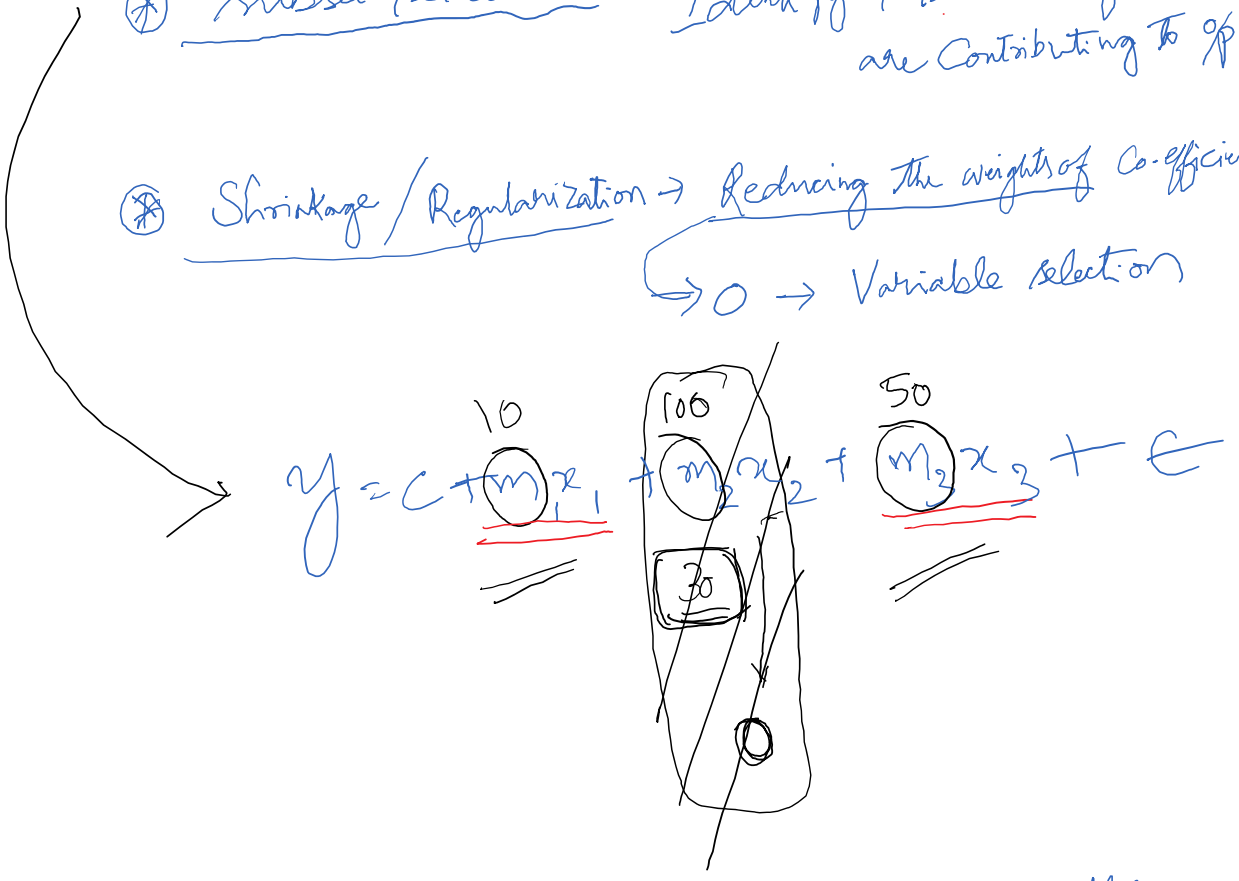
(2) Model Interpritability → Many variables → Not linked to o/p → X

↳ lead to Complexity

* Subset selection → Identify & select useful variables that are Contributing to o/p.

* Shrinkage / Regularization → Reducing the weights of Co-efficients

↳ 0 → Variable selection



$$y = c + m_1 x_1 + m_2 x_2 + m_3 x_3 + \epsilon$$

* Dimensionality Reduction → Unknowns variables → Projections

Linear combination ← x_1 x_2 x_3 x_4 x_5

Subset selection

→ Choose the best model

→ C_p, AIC, BIC & Adj R^2

→ Validation & Cross Validation

Shrinkage methods

→ Ridge Regression

→ Lasso Regression

→ Elastic Net

Dimensionality Reduction

→ Principle Component Analysis → PCA.

→ Partial Least Square