Variance :-

→ 1. refers to the amount by which the prediction of our model will change if we used a diff-diff training set (or test)

2. Prediction for a given point vary b/w diff realization of the model

unseen data point (test data) — Expected mean after giving diff-diff sample on Model.

$Var(f'(x)) = E\left[\left(f'(x) - E[f'(x)]\right)^2\right]$

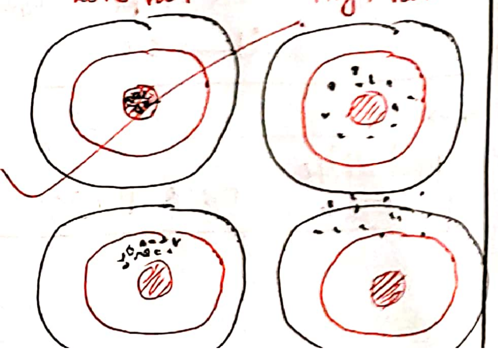
It tells on an avg by changing diff data point what is change in expected value of our model.

st came from $Var = \frac{\sum (x_i - \bar{x})^2}{n}$

$Bias(f(x)) = E[f'(x)] - f(x)$

$Variance(f(x)) = E\left[\left(f(x) - E[f(x)]\right)^2\right]$

Low var   High var



* Bias variance decomposition

$\boxed{Loss = bias + Variance + \text{irreducible error.}}$

↓ further simplified as

$\boxed{Loss = \underbrace{bias^2 + variance}_{\text{reducible Error}} + \underbrace{Var(\epsilon)}_{\text{irr. Error}}}$

Let's

x_1 x_2 ...	y	$\hat{y}$	Error
	8	9	1
	8	8.1	0.1
	7	6.9	0.1
	9	10.1	1.1

Error → reducible, irreducible.

Let's assume (assumption)

$mean_{irreducible} = 0$

$Variance_{irreducible} = \sigma^2$ (const) — spread of irreducible error

Let's take Loss function as MSE

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

$= E[(y-\hat{y})^2]$ {we can write as in terms of expected val.}

$y = \underbrace{f(x)}_{\text{true fun}} + \underbrace{\epsilon}_{\text{Error}} \Rightarrow \theta + \epsilon$

$\hat{y} = f'(x) \Rightarrow \hat{\theta}$

$Loss = E[(y-\hat{y})^2]$

$\Rightarrow E[((\theta+\epsilon) - \hat{\theta})^2]$ After replacing.

$\Rightarrow E[(\underbrace{(\theta-\hat{\theta})}_{a} + \underbrace{\epsilon}_{b})^2]$

$= E[(\theta-\hat{\theta})^2 + \epsilon^2 + 2\epsilon(\theta-\hat{\theta})]$

$[(a+b)^2 = a^2+b^2+2ab]$

$= E[(\theta-\hat{\theta})^2] + E[\epsilon^2] + E[2\epsilon(\theta-\hat{\theta})]$

$[\because E(x+y) = E(x)+E(y)]$

$= E[(\theta-\hat{\theta})^2] + E[\epsilon^2] + E[2] \cdot E[\epsilon] E[\theta-\hat{\theta}]$

$\because E[xy] = E[x]+E[y]$

$= E[(\theta-\hat{\theta})^2] + E[\epsilon^2] + E[2] \overset{0}{E[\epsilon]} E[\theta-\hat{\theta}]$

∵ as we assume Mean of error = 0 & it's a assumption under linear regression.

$= E[(\theta-\hat{\theta})^2] + \underline{E[\epsilon^2]}$

Let's discuss $E[\epsilon^2]$

$Var(\epsilon) = \sigma^2 = E[(\epsilon - \overset{0}{E[\epsilon]})^2]$ — 0 ∵ its assumption

$= E[(\epsilon - 0)^2] = E[\epsilon^2]$

$\boxed{\therefore E(\epsilon^2) = Var(\epsilon)}$

$\therefore MSE = E[(\theta-\hat{\theta})^2] + \underbrace{Var(\epsilon)}_{\text{irreducible}}$

Let's expand this

$E[(\theta-\hat{\theta})^2] = E[(\underbrace{\theta - E[\hat{\theta}]}_{a} + \underbrace{E[\hat{\theta}] - \hat{\theta}}_{b})^2]$

$= E[(\theta-E[\hat{\theta}])^2 + (E[\hat{\theta}]-\hat{\theta})^2 + 2(\theta-E[\hat{\theta}])(E[\hat{\theta}]-\hat{\theta})]$

$= E[(\theta-E[\hat{\theta}])^2] + E[(E[\hat{\theta}]-\hat{\theta})^2] + E[2(\theta-E[\hat{\theta}])(E[\hat{\theta}]-\hat{\theta})]$

$= E[(\theta-E[\hat{\theta}])^2] + E[(E[\hat{\theta}]-\hat{\theta})^2] + \underbrace{E[2]}_{const} \cdot \underbrace{E[(\theta-E[\hat{\theta}])]}_{const} \cdot E[(E[\hat{\theta}]-\hat{\theta})]$

$= E[(\theta-E[\hat{\theta}])^2] + E[(E[\hat{\theta}]-\hat{\theta})^2] + 2(\theta-E[\hat{\theta}])\{\overset{0}{E[E[\hat{\theta}]] - E[\hat{\theta}]}\}$

$\because E[E[\hat{\theta}]] = E[\hat{\theta}]$

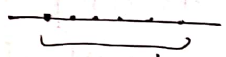
$= E[(\theta-E[\hat{\theta}])^2] + \underline{E[(E[\hat{\theta}]-\hat{\theta})^2]}$

∴ Constant? $\because y = \theta + \epsilon$ → constant
θ → true function i.e. constant
const − const ⇒ const

$= (\theta - E[\hat{\theta}])^2 +$ variance

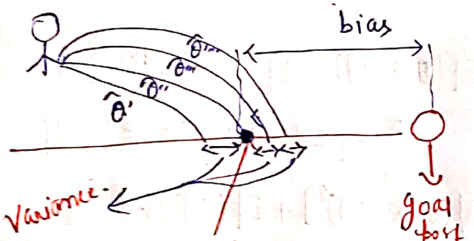
↓ bias² +

$MSE = \underbrace{(bias)^2 + variance}_{\text{reducible Error}} + \underbrace{Var(\epsilon)}_{\text{irreducible error}}$



spread of irreducible error.

Let's assume No irreducible error or noise.



mean of all $\hat{\theta}', \hat{\theta}'', \hat{\theta}''', \hat{\theta}''''$
↓
$E[\hat{\theta}]$

goal post $\theta = f(x)$ ↓ actual

$[E(\hat{\theta}) - \theta] \rightarrow$ bias

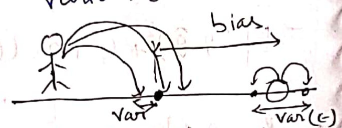
One error; is jo goal keeper shot ka ~~Variance~~ mean tha wo kitna actual se door tha
↓
bias.

2nd Error

↳ Wahi goal keeper ne shots khele ~~wo~~ wo shots ~~sare shots~~ sare shots ke mean (x) kitha distance pe aaT.

⇓

Variance.



Var(ϵ) ?
↳ If our goal post is keep on changing.
variance of that change is called Var(ϵ)

$\boxed{MSE = (bias)^2 + variance + Var(\epsilon)}$

This is because of loss function.

If we reduce Noise = 0 which is impossible

If loss function ~~xx~~ get's changed bias & variance decomposition changes it's ~~xxx~~ formula but it will have bias and variance term.

***

To Reduce Variance

↳ regularization will be used.

To Reduce bias

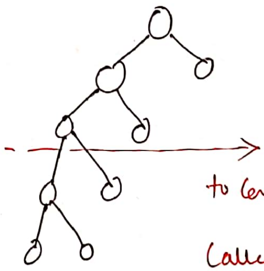
↳ using Complex model.
poly. regr, DT, RF, GB

*** Regularization ?

↳ Technique used to reduce overfitting (variance) in complex model

Regularization technique on diff Model

- Linear model ↓ [Loss function + Extra term] ↓ regularization
- DT ↓ restrict to certain level ⇓ Called as pruning tree.
- Deep Learning ↓ we reduce ~~its~~ nodes in NN ⇓ ~~***~~ Called as Drop out.



* When to use Regularization.

1. Preventing overfitting
2. High Dimensionality :-
   ↳ high # feature compared to # data points. in that scenario, model tends to overfit easily.

   500x40
   200

3. Multicollinearity :-
   ↳ It destabilize model & make the model's estimates sensitive to minor changes in the model.
   (L2) → regularization :- help in such cases by distributing the co-efficient estimates among co-related features.
4. Feature Selection
   ↳ (L1) Lasso :- tends to produce sparse solution, driving the co-eff of irrelevant feature to zero.
5. interpretability
   ↳ regularization can help achieve this by constraining the model's complexity
6. Model Performance
   ↳ It might still improve our model's out-of-sample prediction performance

Scanned with OKEN Scanner