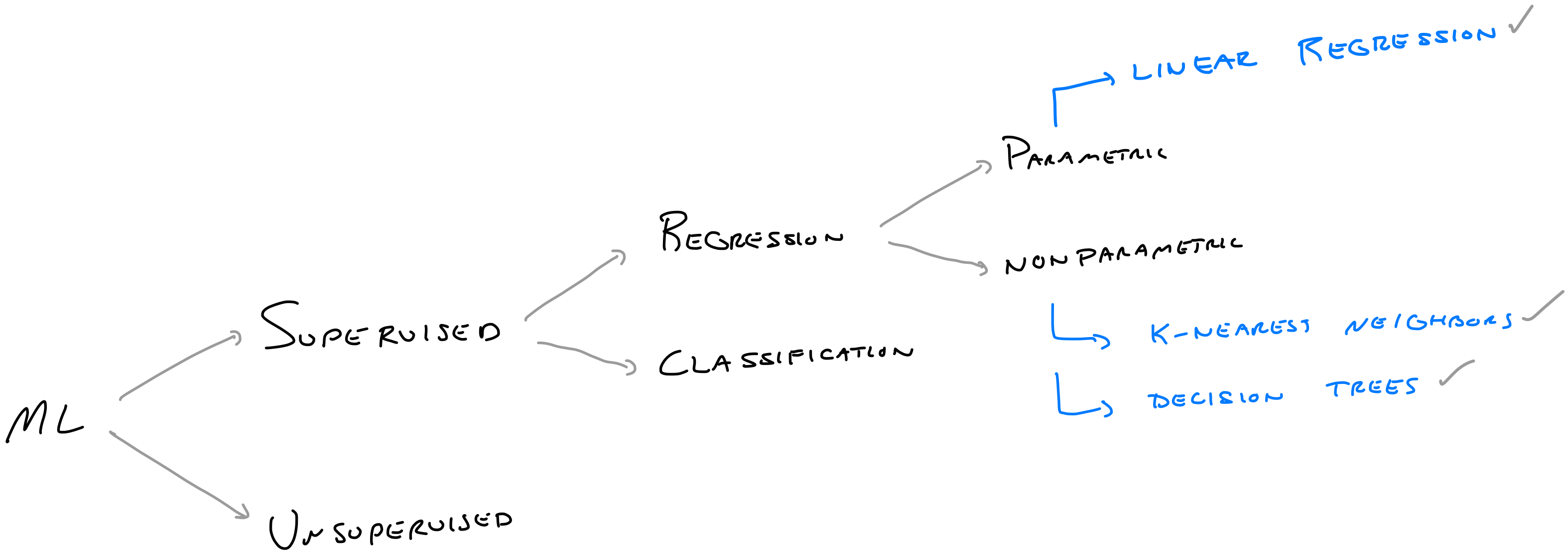
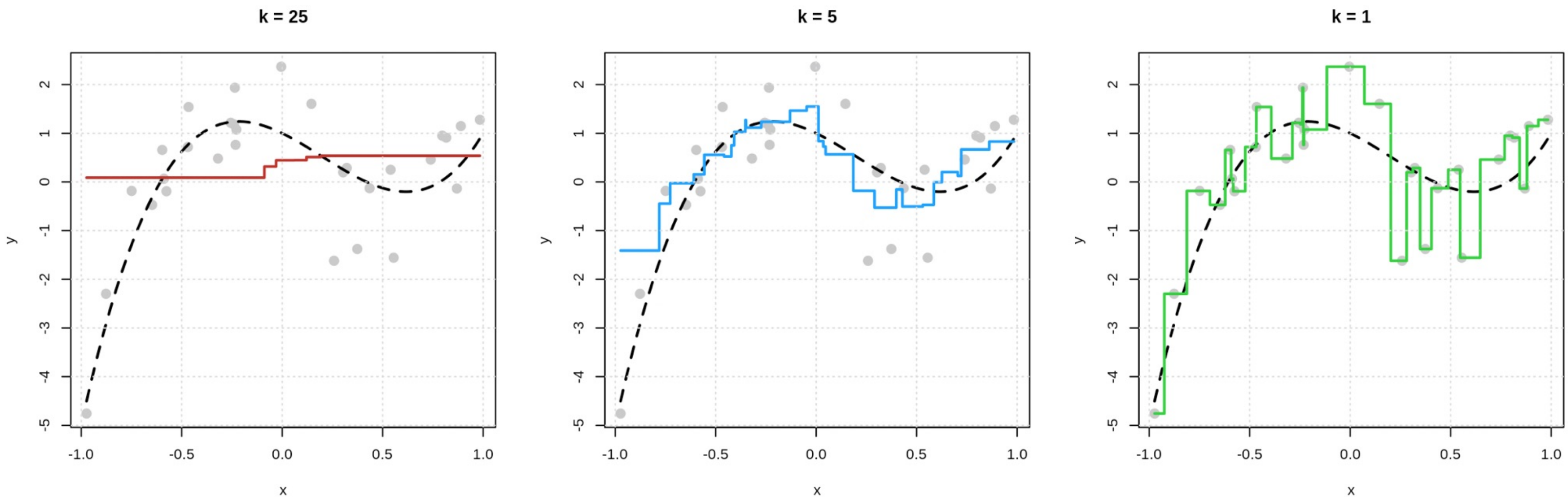
CS 307

Fall 2023

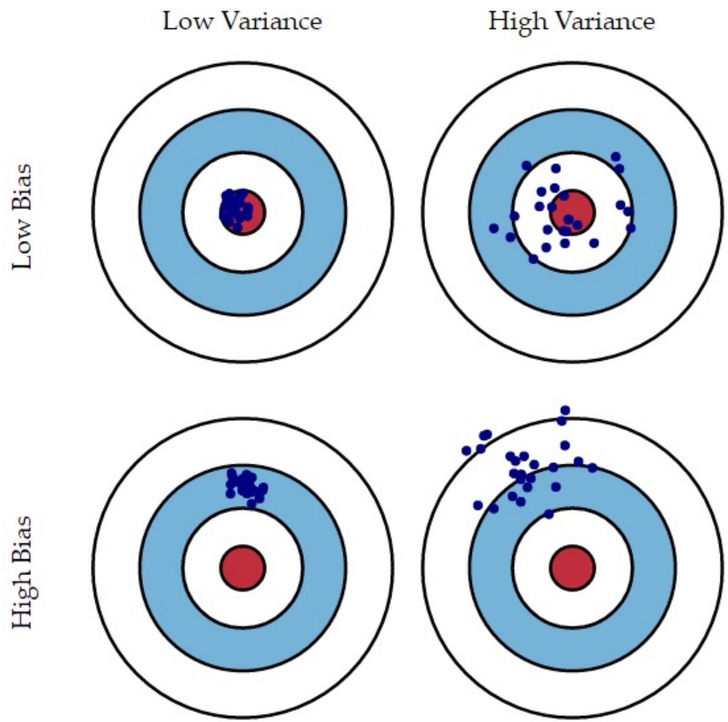
Dalpiaz

$\boxed{\text{Week 04}}$


ML
Supervised
Unsupervised
Regression
Classification
Parametric
Nonparametric
Linear Regression
K-Nearest Neighbors
Decision Trees



k = 25
k = 5
k = 1
y
x
-1.0
-0.5
0.0
0.5
1.0
2
1
0
-1
-2
-3
-4
-5


Low Variance
High Variance
Low Bias
High Bias

θ PARAMETER

$X_1, X_2, \ldots, X_n$ "RANDOM SAMPLE"

$\hat{\theta}(X_1, X_2, \ldots, X_n)$ STATISTIC "ESTIMATOR"

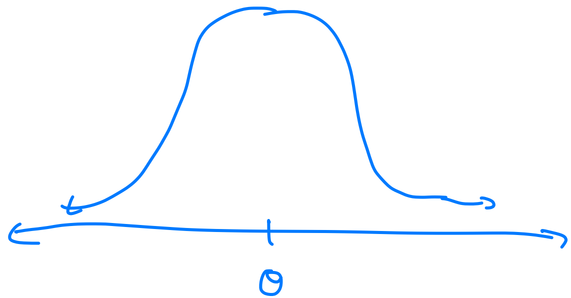
$\hat{\theta} \xrightarrow{\text{ESTIMATE}} \theta$

How good is $\hat{\theta}$ as an estimator of θ?

$\mu \xleftarrow{\text{ESTIMATES}} \hat{\mu}$

$\sigma \xleftarrow{\text{ESTIMATES}} \hat{\sigma}$

θ = MEAN GPA OF UIUC UGs



MODEL ASSUMPTION

"ALL MODELS ARE WRONG,
SOME ARE USEFUL"

– BOX

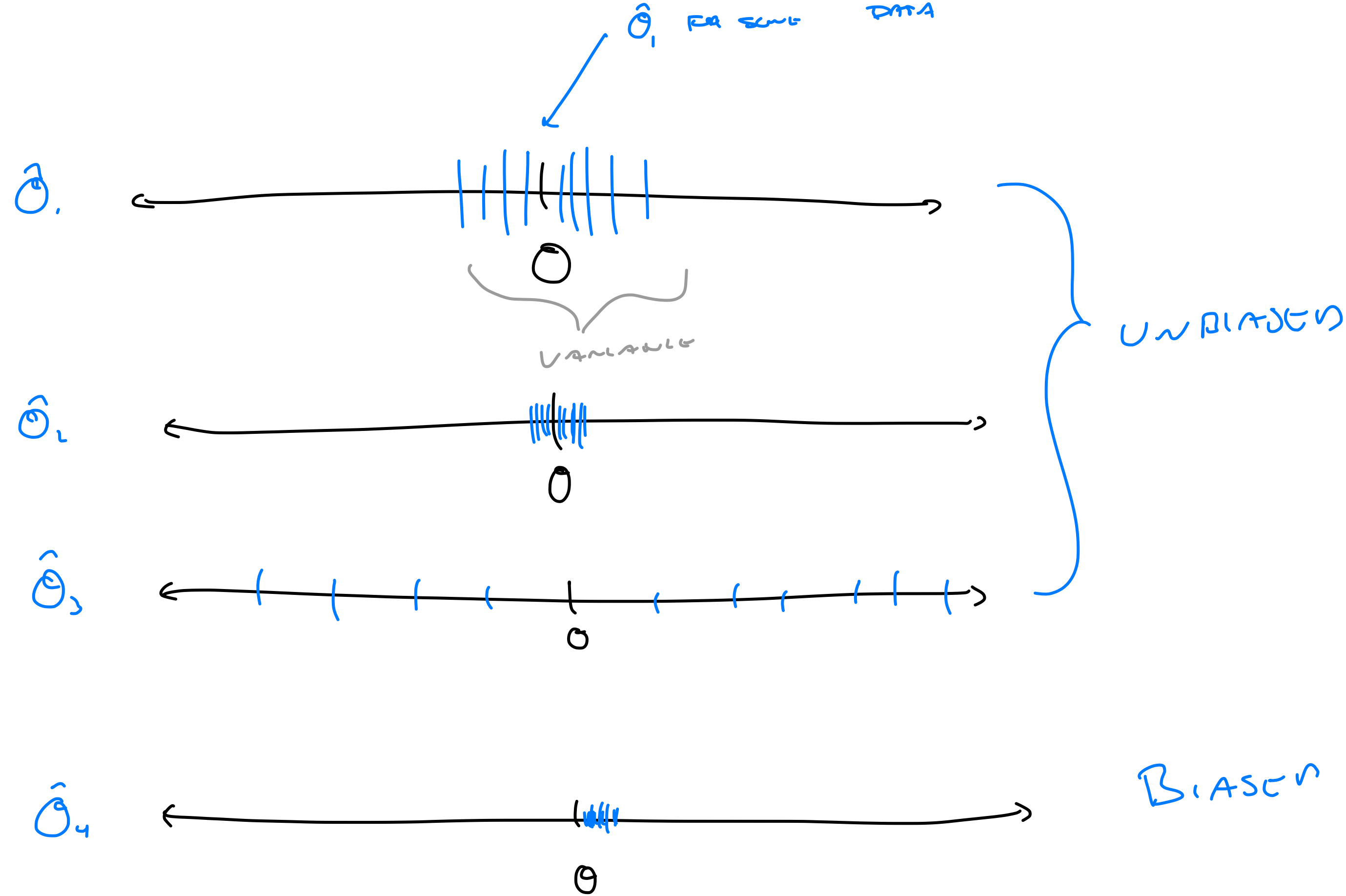
$X_1, X_2, \ldots, X_{30}$ ← RANDOM SAMPLE

$$\hat{\Theta}(\underbrace{X_1, \ldots, X_n})$$

IS GOOD IS $\hat{\Theta}$ AN EST OF Θ?

$\hat{\Theta}$ "CLOSE" Θ → FIXED

RANDOM


$\hat{\theta}_1$ FOR SOME DATA
$\hat{\theta}_1$
θ
VARIANCE
$\hat{\theta}_2$
θ
$\hat{\theta}_3$
θ
UNBIASED
$\hat{\theta}_4$
θ
BIASED

WE WANT $\hat{\theta}$ CLOSE TO θ

"DEFINED AS"

LOSS $\quad L(\theta, \hat{\theta}) \triangleq (\underset{\text{FIXED}}{\theta} - \underset{\text{RANDOM}}{\hat{\theta}})^2$

RISK $\quad R(\theta, \hat{\theta}) = \mathbb{E}[L(\theta, \hat{\theta})] = \mathbb{E}[(\theta - \hat{\theta})^2]$

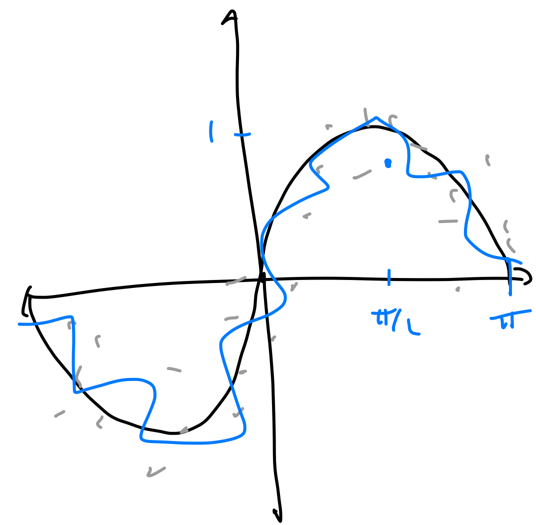
$E[\hat{\theta}]$ MINIMIZES

Data Generating Process

$$Y = \underline{\sin(X)} + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$X \sim U(-\pi, \pi)$$



Y when $x = \pi/2$

DATA

$$(X, Y) \in \mathbb{R}^p \times \mathbb{R}$$

x	y

Assume $Y = f(X) + \varepsilon$

LEARN THIS

f "TRUE UNKNOWN"

$\hat{f}$ LEARN FROM DATA

$$D = \{(x_i, y_i)\}$$

$$L(Y, \hat{f}(x)) = (Y - \hat{f}(x))^2$$

MINIMIZE $$R(Y, \hat{f}(x)) = \mathbb{E}\left[(Y - \hat{f}(x))^2\right]$$

CONDITIONAL MEAN

$$f(x) = \mathbb{E}[Y \mid X = x]$$

$$= \sin(x)$$

TRAINED MODEL

$$EPE\left(Y, \hat{f}(x)\right) = \mathbb{E}_{Y,D|x}\left[\left(Y - \hat{f}(x)\right)^2 \mid X = x\right]$$

EXPECTED PREDICTION ERROR

RANDOM Y

KNOWN x

MINIMIZE

MSE

$$= \underbrace{\mathbb{E}_D\left[\left(f(x) - \hat{f}(x)\right)^2\right]}_{\text{REDUCIBLE ERROR}} + \underbrace{\mathrm{Var}_{Y|x}(Y)}_{\text{IRREDUCIBLE ERROR}}$$

$$\mathrm{MSE}\left(f(x), \hat{f}(x)\right) = \mathbb{E}\left[\left(f(x) - \hat{f}(x)\right)^2\right]$$

$$= \underbrace{\left(f(x) - \mathbb{E}\left[\hat{f}(x)\right]\right)^2}_{\text{BIAS}} + \underbrace{\mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right)^2\right]}_{\text{VARIANCE}}$$

$$\text{Bias}\left[\hat{\theta}\right] = \theta - E\left[\hat{\theta}\right]$$

$$\text{Var}(\hat{\theta}) = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right]$$

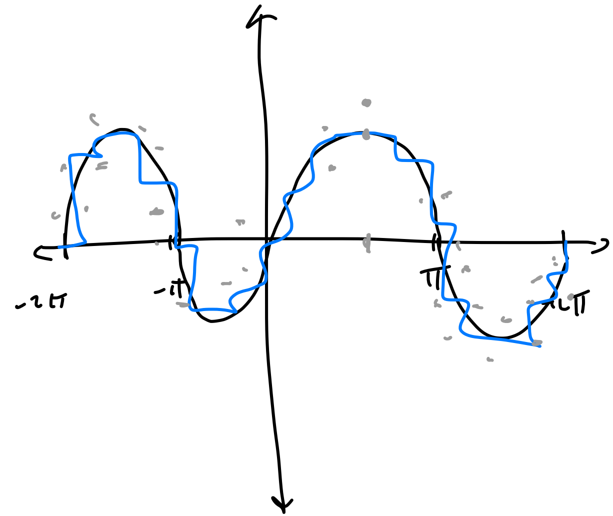
Data Generating Processes

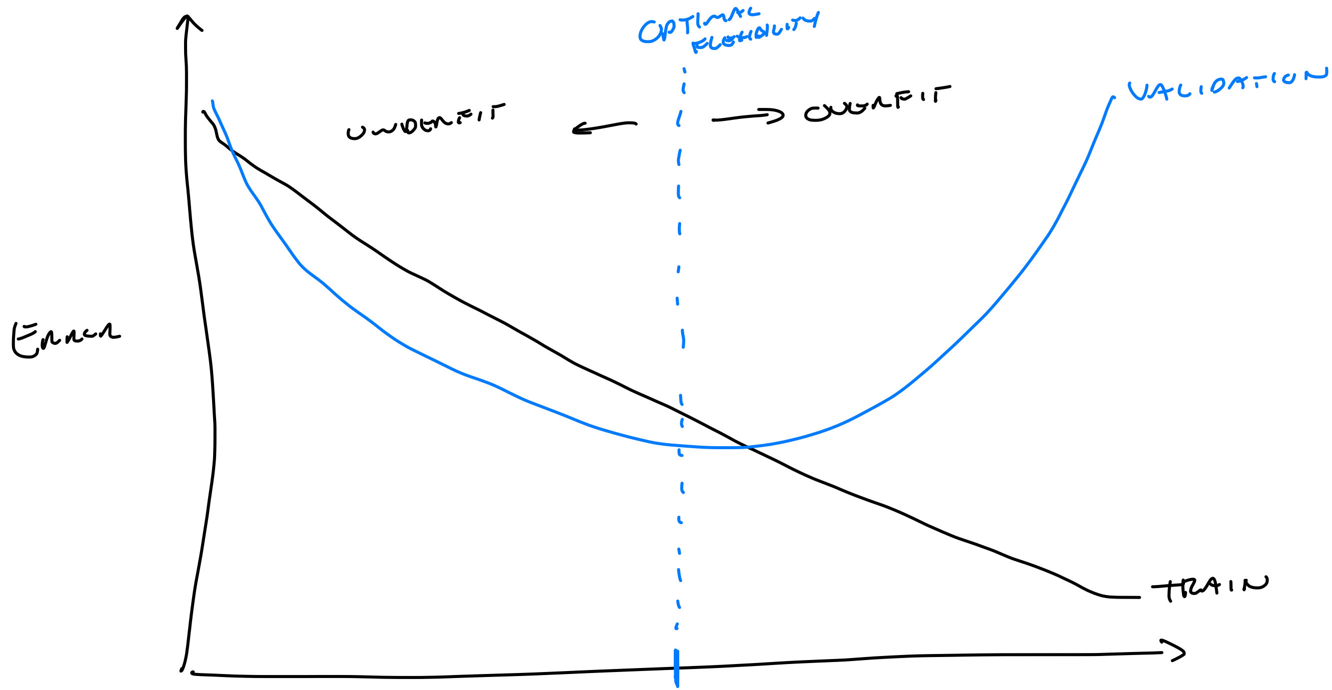
$\rightarrow$ $X \sim U(-2\pi, 2\pi)$

$\varepsilon \sim N(0, \sigma^2)$

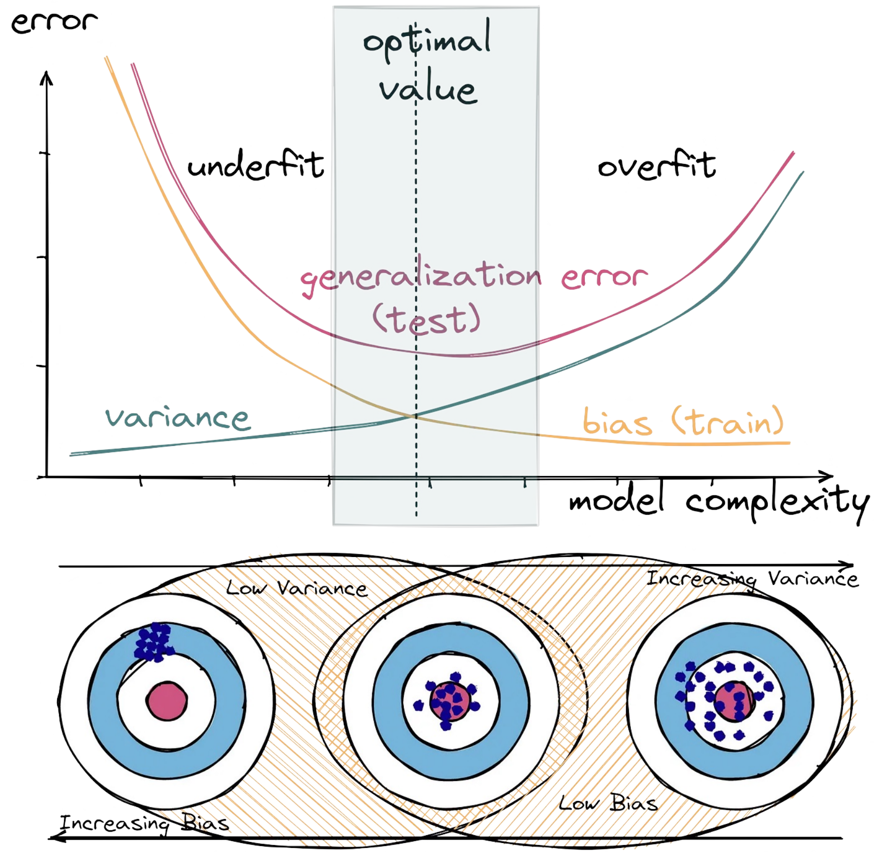
$$Y = \underbrace{\sin(X)}_{\text{SIGNAL}} + \varepsilon$$

NOISE





	LOW	←	FLEXIBILITY	→	HIGH
KNN	HIGH	←	K	→	LOW
	HIGH	←	BIAS	→	DOWN
	LOW	←	VAR	→	HIGH
TREG	HIGH	←	MSS	→	LOW


error
optimal
value
underfit
overfit
generalization error
(test)
variance
bias (train)
model complexity
Low Variance
Increasing Variance
Low Bias
Increasing Bias