

Machine Learning

GO Classes Host

Bias-Variance Tradeoff

www.goclasses.in

00:43 - 1:19:32

Machine Learning

GO Classes Host

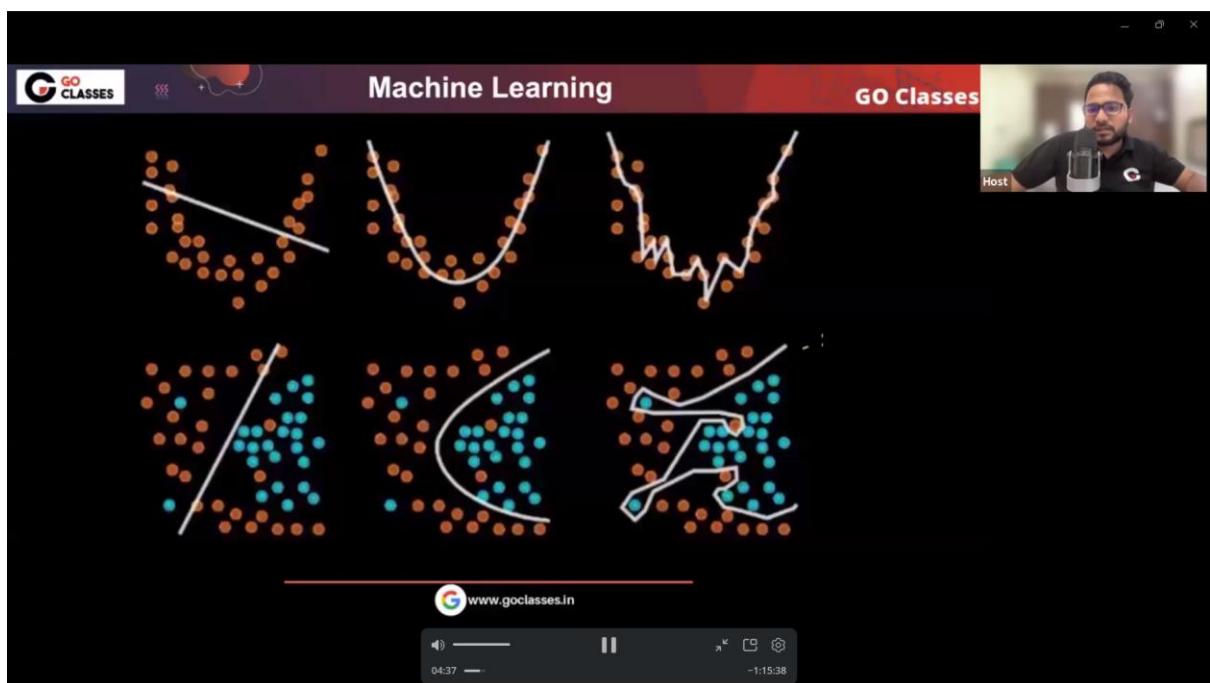
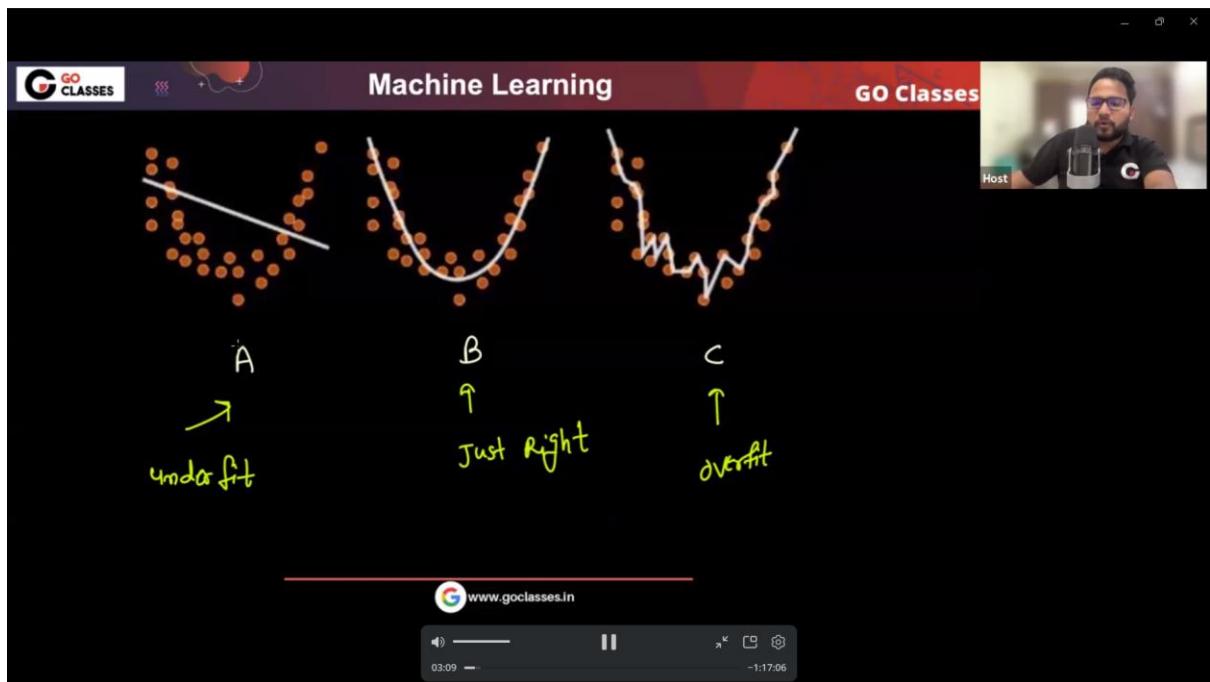
Problem solving approach:

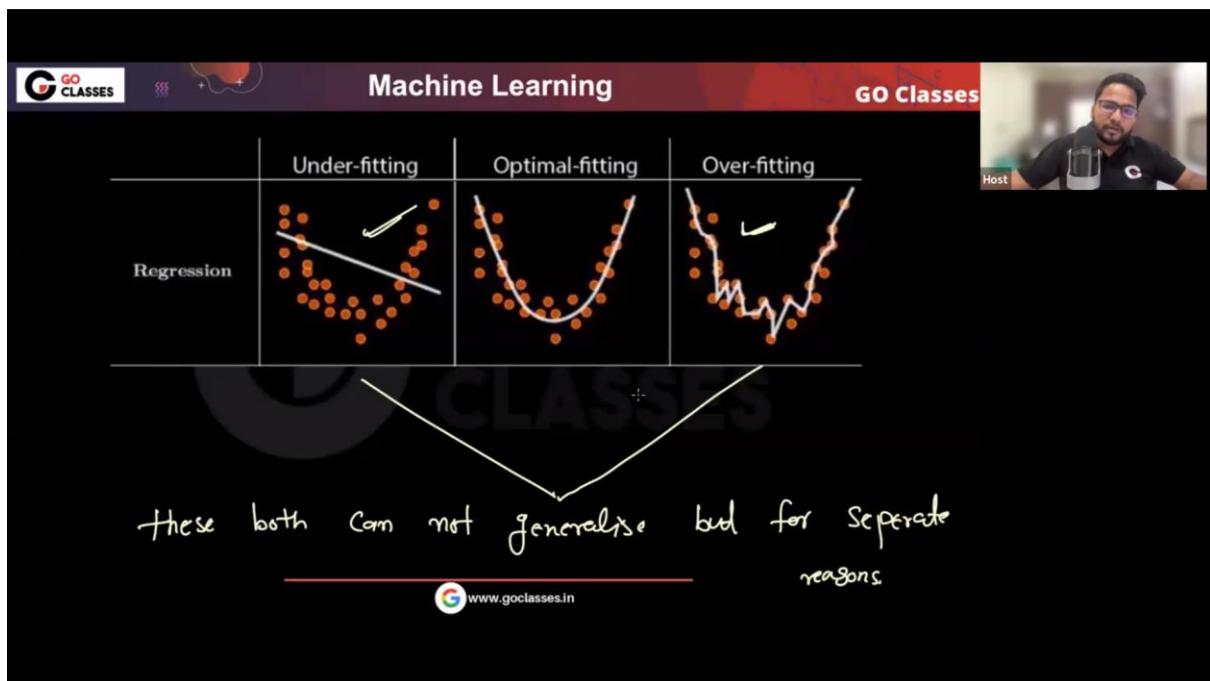
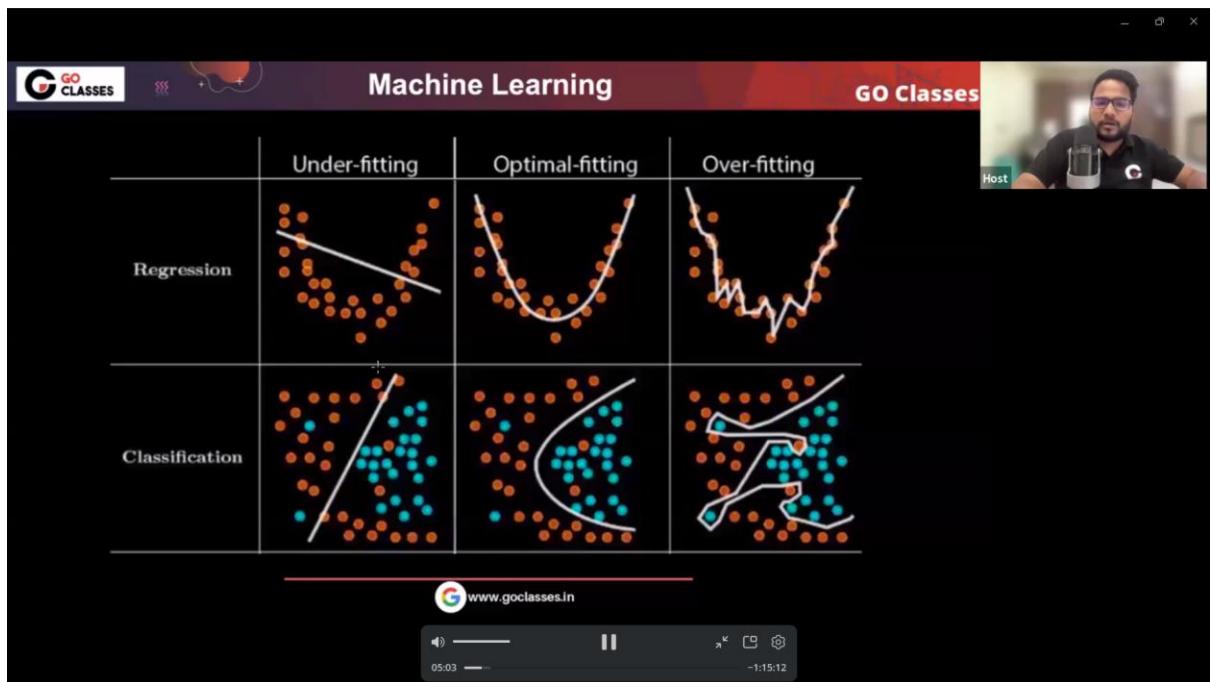
underfitting (simple model)

overfitting

just right

02:02 - 1:18:13





The screenshot shows a video player interface for a Go Classes video. The title bar reads "Machine Learning" and "Generalization error Components". A small video thumbnail of the host is visible in the top right. The main content area displays two bullet points about generalization error components:

- Bias: Component of generalization error due to model is less expressive
- Variance: Component of generalization error due to model is too much expressive and data is less.

At the bottom of the screen, there is a navigation bar with icons for volume, full screen, and other controls, along with the URL "www.goclasses.in" and the current timestamp "09:42 - 1:10:33".

The screenshot shows a video player interface for a Go Classes video. The title bar reads "Machine Learning" and "Generalization error Components". A small video thumbnail of the host is visible in the top right. The main content area displays two bullet points, each preceded by a yellow checkmark:

- ✗ • Bias: Component of generalization error due to model is less expressive
- ✓ • Variance: Component of generalization error due to model is too much expressive and data is less.

Overlaid on the video content is a handwritten mathematical equation: $\text{Test error} = \underbrace{\underline{x}}_{\text{Bias}} + \underbrace{\underline{y}}_{\text{Variance}}$.

At the bottom of the screen, there is a navigation bar with icons for volume, full screen, and other controls, along with the URL "www.goclasses.in" and the current timestamp "11:11 - 1:09:04".

Bias → underfitting

Variance → overfitting

Machine Learning

The graph illustrates a true model $f(x)$ represented by a smooth black curve, and a complex model $\hat{f}(x)$ represented by a jagged blue line that closely follows the data points (marked with 'x').

- Let $f(x)$ be the true model (usually unknown)
- $MSE = \sum_i (y_i - \hat{y}_i)^2$
- $\hat{y} = \hat{f}(x)$ \hat{f} could be simple or complex model

$\hat{f}(x)$ is complex model in this case.

www.goclasses.in

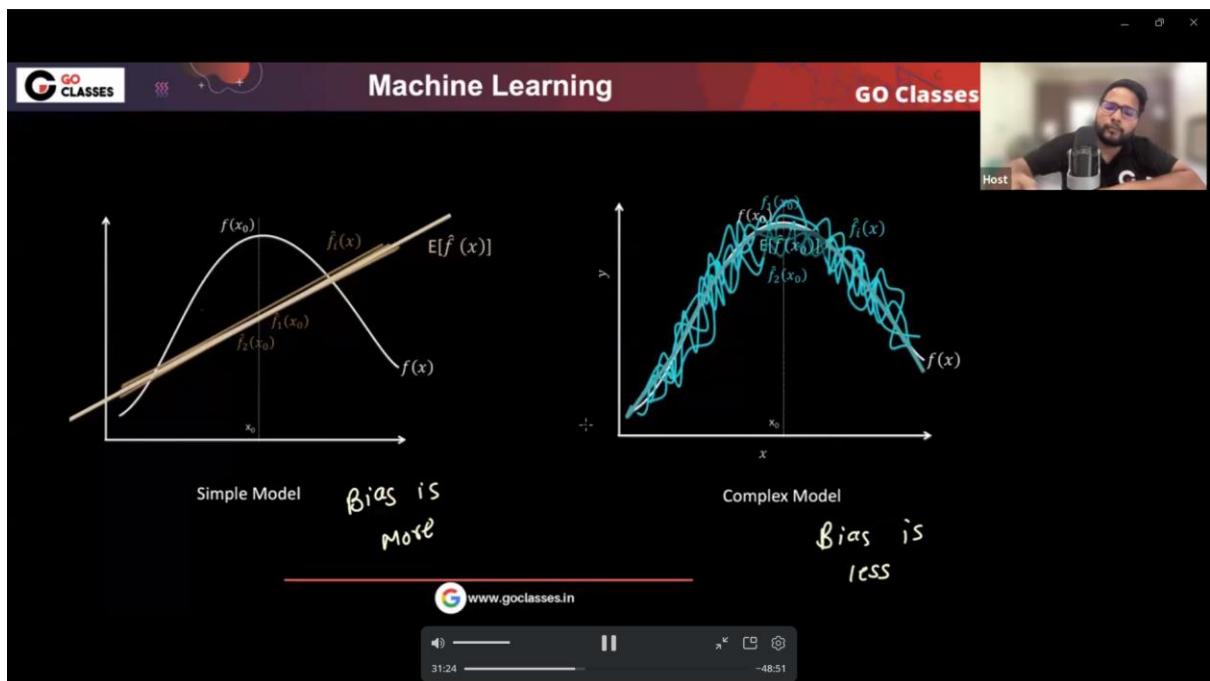
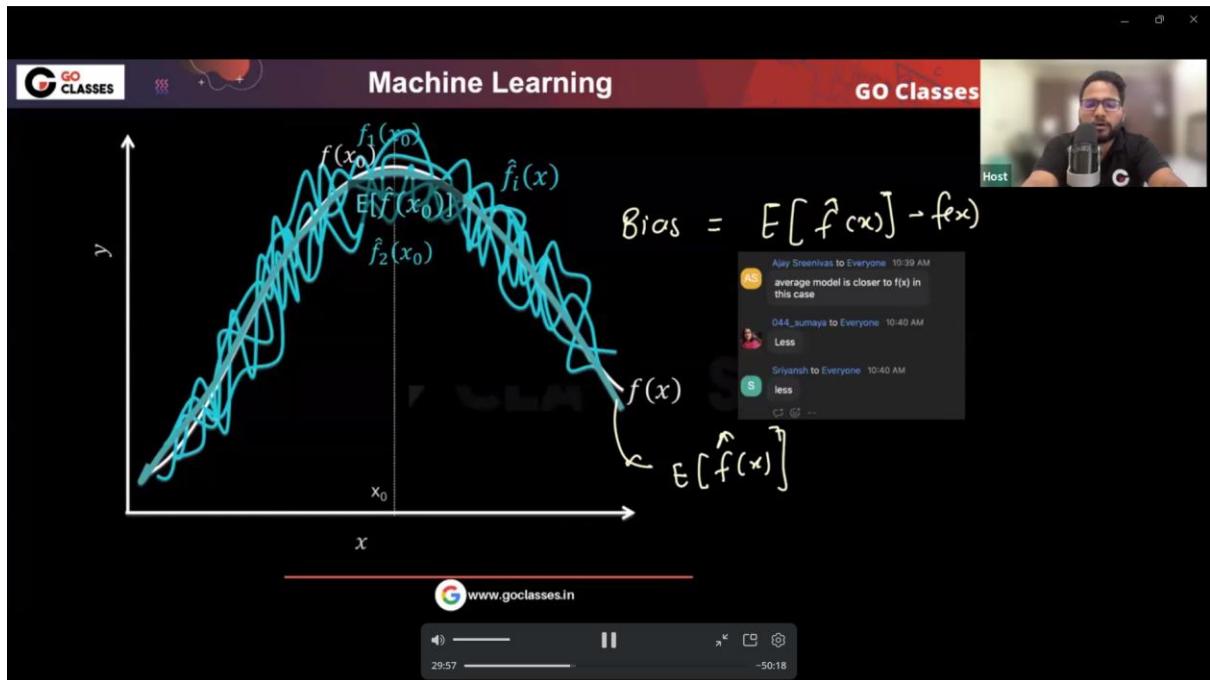
14:18 - 1:05:57

The graph shows a true model $f(x)$ and several estimated models $\hat{f}_1(x)$, $\hat{f}_2(x)$, and $E[\hat{f}(x)]$. The point x_0 is marked on the x-axis. The text indicates that we can keep taking different datasets and getting new models.

We can keep on taking different datasets and keep on getting new models.

$\text{Bias} = E[\hat{f}(x)] - f(x)$ } won't this term will be higher for simple model?

28:12 - 52:03



Bias

- Let $f(x)$ be the true model and $\hat{f}(x)$ be our estimate of the model.

$$\text{Bias}(\hat{f}(x)) = E[\hat{f}(x)] - f(x)$$

Mathematically, this means that the simple model has a high bias

www.goclasses.in

E → estimated

Variance

- Let $f(x)$ be the true model and $\hat{f}(x)$ be our estimate of the model.

$$\text{Variance}(\hat{f}(x)) = E[(\hat{f}(x) - E[\hat{f}(x)])^2]$$
$$a^2 + b^2 + c^2 + x^2 + y^2 + z^2$$

www.goclasses.in

36:11 -44:04

Machine Learning

Host

- In summary (informally)
- Simple model: high bias, low variance
- Complex model: low bias, high variance
- There is always a trade-off between the bias and variance
- Both bias and variance contribute to the mean square error.

www.goclasses.in

Bias → how far it is from true function

Machine Learning

Host

[1 Pt] Which of the following plots depicts models with the highest model variance?

(a) (b) (c)

(a) (b) (c)

www.goclasses.in

41:43 -38.32

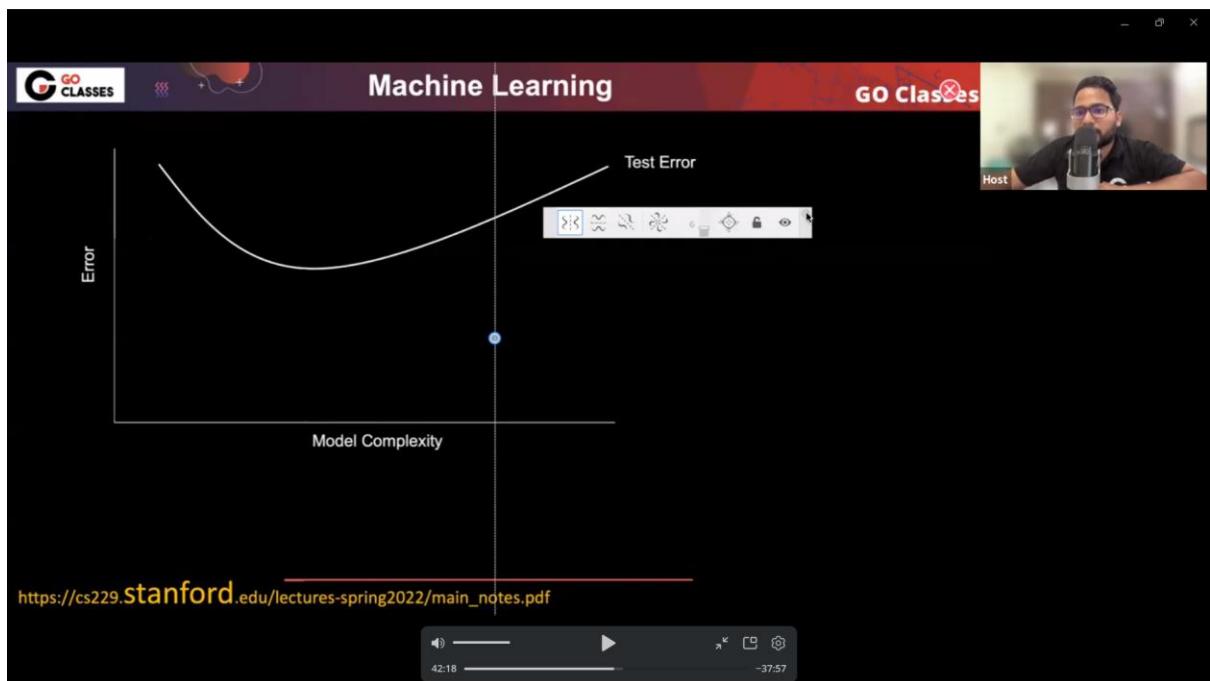
[2 Pts] Which of the following depicts models with the largest bias?

The first plot is selected.

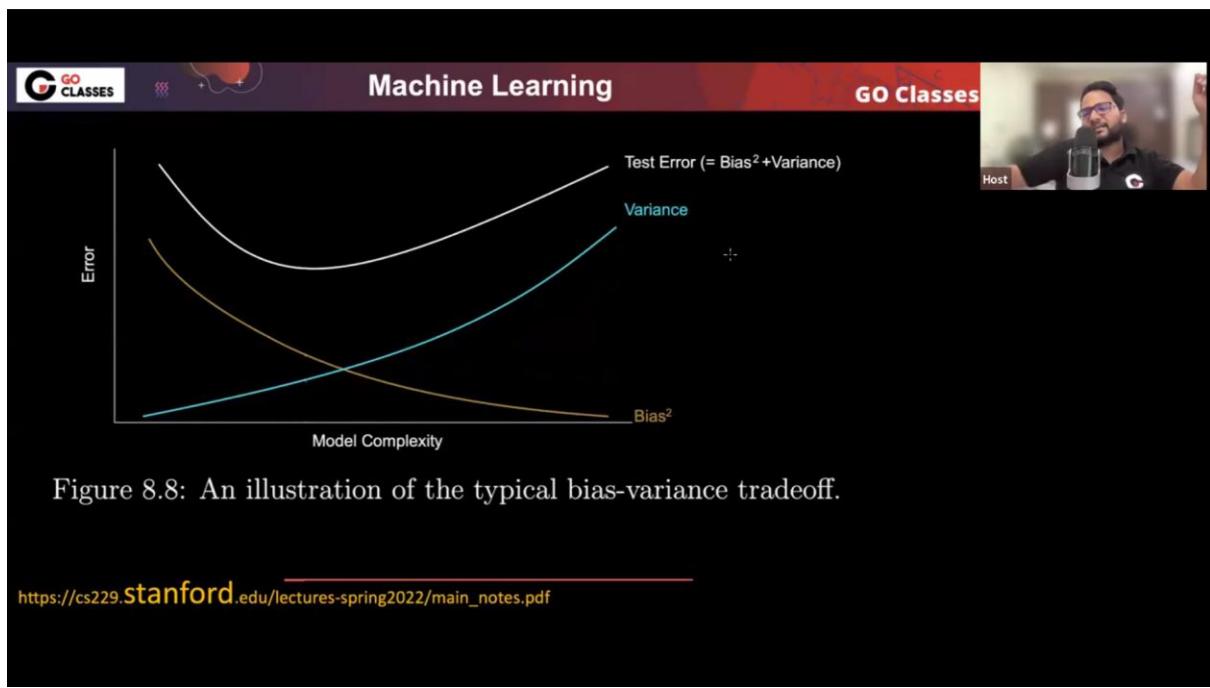
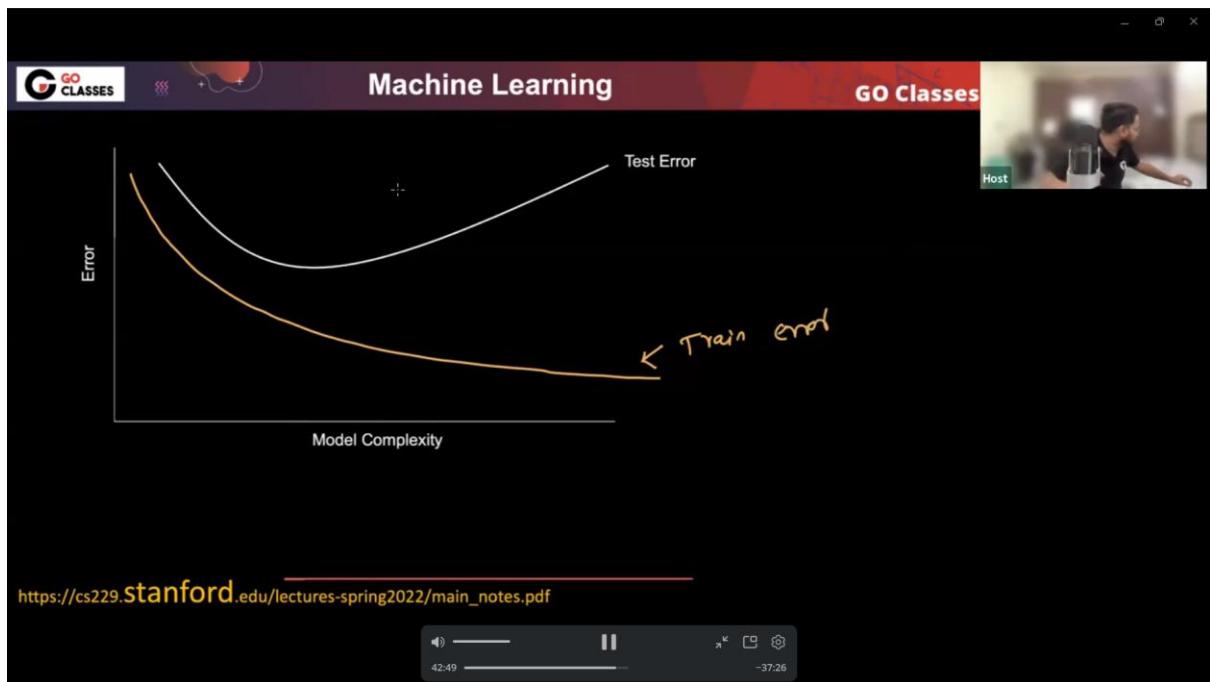
Host

www.goclasses.in

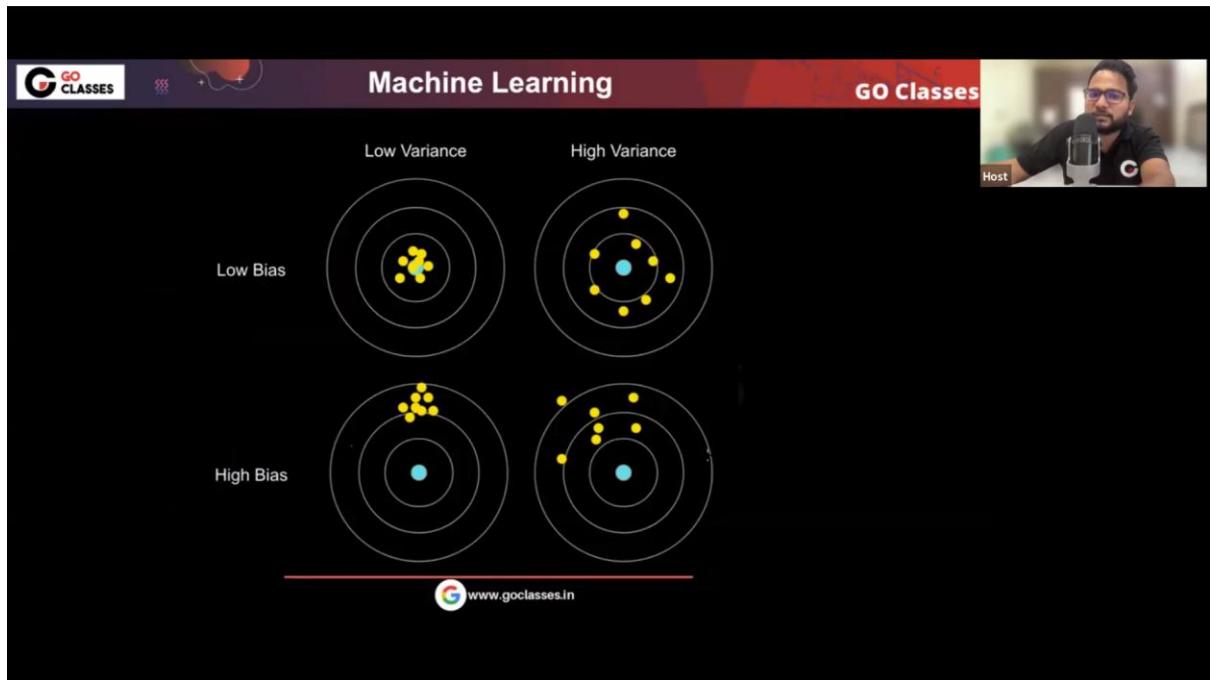
42:05 - 38:10



If we keep on increasing the complexity at some point the test error will also increase..



Predictions distance less → less variance



Question:

I. II. III. IV.

Figure 1: The following graphic will be used as a representation of bias and variance. Imagine that a true/correct model is one that always predicts a location at the center of each target (being farther away from the center of the target indicates that a model's predictions are worse). We retrain a model multiple times, and make a prediction with each trained model. For each of the targets, determine whether the bias and variance is low or high with respect to the true model.

In Figure 1, subplot I, how are bias and variance related to the true model?

- (A) High bias, High variance
- (B) High bias, Low variance
- (C) Low bias, High variance
- (D) Low bias, Low variance

https://courses.cs.washington.edu/courses/cse446/73su/exams/pastexams/72su-midterm-solutions.pdf

51:21 -28:54

Machine Learning

Question:

I. II. III. IV.

In Figure 1, subplot II, now are bias and variance related to the true model?

(A) High bias, High variance
 (B) High bias, Low variance
 (C) Low bias, High variance
 (D) Low bias, Low variance

https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/22au-midterm-solutions.pdf

51:55 -28:20

Host

Machine Learning

Question:

I. II. III. $\equiv B$ IV. \equiv

In Figure 1, subplot II, now are bias and variance related to the true model?

(A) High bias, High variance
 (B) High bias, Low variance
 (C) Low bias, High variance
 (D) Low bias, Low variance

https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/22au-midterm-solutions.pdf

52:07 -28:08

Host

Machine Learning

Question:

The figure shows four targets (circles) with a bullseye in the center.
 I. A single dot at the center.
 II. Many dots tightly clustered around the center.
 III. Many dots scattered widely from the center.
 IV. A few dots scattered widely from the center.

III. $\equiv B$ IV. $\equiv A$

In Figure 1, subplot II, how are bias and variance related to the true model?

(A) High bias, High variance
 (B) High bias, Low variance
 (C) Low bias, High variance
 (D) Low bias, Low variance

https://courses.cs.washington.edu/courses/cse446/23au/exams/pastexams/22au-midterm-solutions.pdf

52:18 -27:57

Host

Question:

[1 points] What best defines the relationship between a model's fit and its bias and variance

(a) A model with low bias and high variance is underfitting
 (b) A model with high bias and high variance is underfitting
 (c) A model with low bias and low variance is underfitting
 (d) A model with high bias and low variance is underfitting

The graph plots Error (Y-axis) against Complexity (X-axis).
 - The 'Training Error' curve decreases monotonically as complexity increases.
 - The 'Test Error' curve starts high, decreases to a minimum (labeled 'Getting Better'), and then increases sharply towards infinity as complexity continues to increase.
 - The region where the test error is high and the training error is low is labeled 'Underfitting!'.
 - The region where both errors are very low is labeled 'Overfitting!'.

Underfitting and Overfitting

Error

High Bias Low Variance Low Bias High Variance

Complexity

Training Error

Test Error

Underfitting!

Overfitting!

Getting Better

https://www.seas.upenn.edu/~cis520/exams/midterm2022_solutions.pdf

53:53 -26:22

Host

Machine Learning

[1 points] What best defines the relationship between a model's fit and its bias and variance

(a) A model with low bias and high variance is underfitting
(b) A model with high bias and high variance is underfitting
(c) A model with low bias and low variance is underfitting
(d) A model with high bias and low variance is underfitting

★ SOLUTION: D ✓

https://www.seas.upenn.edu/~cis520/exams/midterm2022_solutions.pdf

https://www.seas.upenn.edu/~cis520/exams/midterm2022_solutions.pdf

Machine Learning

How will regularizing the weights in a linear regression model change the bias and variance (relative to the same model with no regularization)?

(A) Increase bias, increase variance
(B) Increase bias, decrease variance
(C) Decrease bias, increase variance
(D) Decrease bias, decrease variance

Solution:

The solution is (B).

www.goclasses.in

56:02 -24:13

Machine Learning

Question:

How will regularizing the weights in a linear regression model change the bias and variance (relative to the same model with no regularization)?

(A) Increase bias, increase variance
(B) Increase bias, decrease variance
(C) Decrease bias, increase variance
(D) Decrease bias, decrease variance

Regularization
↓
reduce overfitting
↓
reduce variance

✓

<https://courses.cs.washington.edu/courses/cse446/23wi/assignments/midterm/midterm-solutions.pdf>

Machine Learning

(1 point) Peanut wants to train a model to accurately classify different types of animals from images. After training and testing his model, he observes that the model has high training error and high test error. What can we most confidently say about the bias/variance characteristics of Peanut's model?

A. High bias. ✓ → model is simple
B. Low bias.
C. High variance.
D. Low variance. ✓

(because high training error is given)

Solution: A. High bias because the model has high training error.

High training error ⇒ Simple model
⇒ High bias
Low variance

www.goclasses.in

GO CLASSES Machine Learning GO Classes Host

Question: MSQ

Which of the following can impact our model variance? Select all that apply.

- The regularization coefficient λ .
- The choice of features to include in our design matrix.
- The learning rate α in gradient descent.
- The size of the training set.

Variance is reduced if Size of training set increases

↑ Complicated features

↓ complex model

↓ Variance is high

www.goclasses.in
1:06:41 - 13:34

GO CLASSES Machine Learning GO Classes Host

Which of the following are indications that you should regularize? Select all that apply.

- A. Our training loss is 0.
- B. Our model bias is too high.
- C. Our model variance is too high.
- D. Our weights are too large.

www.goclasses.in
1:09:16 - 10:59

Machine Learning

Host

Reducing the regularization of a model would typically . . .

(a) Decrease its bias and increase its variance
(b) Decrease its bias and decrease its variance
(c) Increase its bias and decrease its variance
(d) Increase its bias and increase its variance

Correct answers: (a)

www.goclasses.in

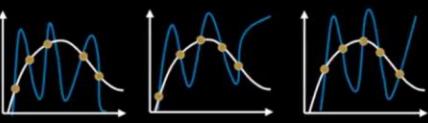
1:16:46 -03:29

Machine Learning

Host

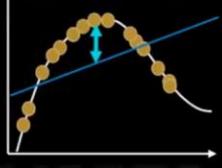
Variance

Regardless of what the true function is, how much does our prediction change with dataset?



Bias²

This error doesn't go away no matter how much data we have!



Variance reduces if we increase data

$$\text{Bias} = E(\hat{f}(x)) - f(x)$$

www.goclasses.in

Machine Learning

Consider a quintic (degree 5) model $\hat{y} = \hat{q}_0 + \hat{q}_1 x + \dots + \hat{q}_5 x^5$, where \hat{q}_j is the MSE parameter estimate for feature j of this quintic model. You first fit this model to Training Set #1; then, you fit this model again to Training Set #2.

(i) The primary contributor to this model's risk is: High bias High variance

(ii) This contributor to model risk can be addressed by the following (select all that apply):

- Reducing model complexity
- Adding additional features to the model
- Reducing the regularization hyperparameter λ
- Increasing the regularization hyperparameter λ

https://ds100.org/sp24/resources/assets/exams/sp23/sp23_final_sol.pdf

Host

16:16 - 1:42:54

(b) [1 Pt] Suppose we control the complexity of the linear models using a Ridge penalty term $\lambda \sum \beta_i^2$. Which of the following is true?

- The left side of the graph represents small λ .
- The right side of the graph represents small λ .

Solution: A smaller λ value means higher model complexity. Remember that a zero λ value means a model with no regularization.

(c) [3 Pts] Which of the following can impact our model variance? Select all that apply.

- The regularization coefficient λ .
- The choice of features to include in our design matrix.
- The learning rate α in gradient descent,
- The size of the training set.

www.goclasses.in

21:03 - 1:38:07