

A1

a

Bias refers to the error introduced by approximating a complex real-world problem with a simpler model. High bias means the model is underfitting the data. Variance is the error introduced by the model's sensitivity to small changes in the training data. High variance indicates overfitting. The bias-variance trade-off involves balancing these two types of errors. As model complexity increases, bias decreases but variance increases, and vice versa. The goal is to find a balance that minimizes total error.

b

When model complexity increases, bias decreases because the model can better fit the training data, but variance increases because the model may overfit. When model complexity decreases, variance decreases because the model becomes less sensitive to the training data, but bias increases because the model might underfit.

c

False. While reducing features can help prevent overfitting and improve generalization, this is not always true. If we remove important features, the model may lose predictive power and underfit instead.

d

False. Hyperparameters should not be tuned on the test set because it would lead to overfitting the test data, making it no longer a reliable measure of generalization performance.

Procedure for Hyperparameter Tuning:

Split the data into training, validation, and test sets. Train the model on the training set and use the validation set to tune hyperparameters. Once the best parameters are selected, evaluate the model on the test set.

e

False. The training error typically underestimates the true error because the model has already seen the training data and is optimized for it. The true error is better estimated using validation or test data.