

Bias & Variance: The ML Balancing Act



What is Bias?

- Bias is the error from overly simplistic assumptions in your model.
- A high-bias model fails to capture the underlying trends in the data.
- It oversimplifies the problem, leading to inaccurate predictions.
- Think of it as an underfitting model.
- It performs poorly on both training and test data.
- Linear regression on complex data is a high-bias example.



What is Variance?

- Variance is the error from sensitivity to small fluctuations in the training set.
- A high-variance model learns the noise in the data, not just the signal.
- It overcomplicates the problem, memorizing the training data.
- Think of it as an overfitting model.
- It performs well on training data but poorly on unseen test data.
- A complex deep learning model with little data is high-variance.



The Trade-Off

- You cannot minimize bias and variance at the same time.
- Reducing bias typically increases variance.
- Reducing variance typically increases bias.
- The goal is to find the optimal balance for your data.
- This is known as the bias-variance trade-off.
- The perfect model has low bias and low variance.



Visualizing The Problem

- Imagine trying to hit the bullseye on a target.
- High Bias: Your shots are consistently off-target but grouped.
- High Variance: Your shots are scattered all around the target.
- High Bias & High Variance: Shots are scattered and off-target.
- Ideal: Your shots are tightly grouped on the bullseye.
- This analogy helps understand model consistency and accuracy.



High Bias Solutions

- Use a more complex algorithm (e.g., trees instead of linear).
- Add more relevant features to the dataset.
- Reduce regularization strength, as it can increase bias.
- Train the model for a longer period.
- The key is to give the model more learning capacity.
- Addresses the problem of underfitting.



High Variance Solutions

- Get more training data (the most effective method).
- Use a less complex algorithm (e.g., linear instead of trees).
- Apply feature selection to reduce redundant inputs.
- Increase regularization strength (e.g., L1, L2).
- Use techniques like pruning for decision trees.
- Addresses the problem of overfitting.



Model Complexity Link

- Model complexity is a key driver of the trade-off.
- As model complexity increases, bias decreases.
- As model complexity increases, variance increases.
- Total error has a sweet spot at a certain complexity level.
- Your job is to find this optimal point for your specific task.
- This is why simple models are often a good starting point.



Key Takeaway

- All supervised learning models have bias and variance.
- Understanding this trade-off is fundamental to building good models.
- Diagnose your model's error to know if it's high bias or high variance.
- Choose strategies that target the specific problem you have.
- The goal is a model that generalizes well to new, unseen data.
- Mastering this balance is a core skill in machine learning.



Follow for More



Like this post? Repost!

Share your thoughts in the comments!

I share little hacks and tips I actually use in
Data & AI — follow if you wanna see more!



Shoaib Akthar
@theshoaibakthar

Follow for more Data Analytics, Data Science and AI insights