

Bias Vs Variance

Parth Bansal

June 2025

Diagnosing High Bias vs. High Variance

To determine whether a model suffers from high **bias** or high **variance**, we compare the model's performance on the training and validation (or test) sets.

Guideline Table

Training Error	Validation Error	Interpretation
High	High	High Bias (Underfitting)
Low	High	High Variance (Overfitting)
Low	Low	Good Fit (Well Generalized)

Practical Diagnosis

- **High Bias (Underfitting):** The model is too simple to capture the underlying pattern in the data. Both training and validation errors are high.
- **High Variance (Overfitting):** The model fits the training data very well but fails to generalize to unseen data. Training error is low, but validation error is high.

Tools and Techniques

- **Learning Curves:** Plot training and validation errors versus training set size to visualize bias and variance behavior.
- **Cross-Validation:** If performance varies significantly across folds, variance is likely high.
- **Model Complexity Check:** Try using simpler or more complex models to observe how errors change.

Code Example (Python-like)

```
from sklearn.metrics import mean_squared_error

train_error = mean_squared_error(y_train, model.predict(X_train))
val_error = mean_squared_error(y_val, model.predict(X_val))

print(f"Train Error: {train_error}, Val Error: {val_error}")
```

Interpret the results based on the earlier table.

How to Fix

High Bias (Underfitting)	High Variance (Overfitting)
Add more features	Get more training data
Use a more complex model	Use a simpler model
Reduce regularization ($\lambda \downarrow$)	Increase regularization ($\lambda \uparrow$)
Train for more epochs	Use dropout / early stopping

What is an Epoch?

An **epoch** is one complete pass through the entire training dataset during the learning process. In each epoch, the model updates its parameters (e.g., weights) by computing gradients and applying updates (e.g., via gradient descent). Training for multiple epochs allows the model to iteratively learn better representations of the data. However, too many epochs can lead to overfitting, while too few may result in underfitting.