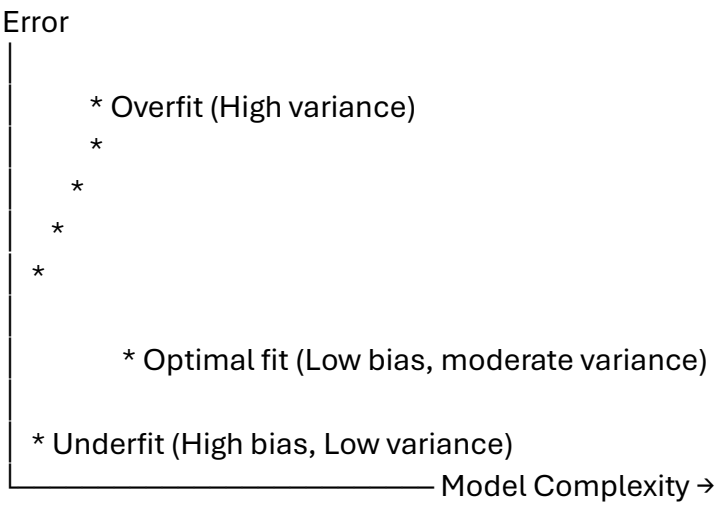
Assignment 01

. Bias and Variance Concepts

- **Bias**: Error due to overly simplistic assumptions in the model.
    - High bias → model is too simple → **underfitting**.
    - Low bias → model can capture the data patterns better.
- **Variance**: Error due to the model being too sensitive to small fluctuations in the training data.
    - High variance → model is too complex → **overfitting**.
    - Low variance → model is stable across different datasets.

2. Typical Diagram (Bias-Variance Tradeoff)



- As **model complexity increases**:
    - Bias decreases (model fits training data better)
    - Variance increases (model is sensitive to small changes in data)
- **Goal:** Minimize total error = $Bias^2$ + Variance + Irreducible error

3. Overfitting vs Underfitting

Problem	Bias	Variance	Model characteristic
Underfitting	High	Low	Too simple, misses patterns

Problem	Bias	Variance	Model characteristic
Overfitting	Low	High	Too complex, fits noise
Good fit	Low	Moderate	Captures pattern, generalizes

4. Best-Fit Model

- **Best fit = low bias + low variance** (as much as possible)
- Practically, there's always a tradeoff:
    - Extremely low bias → can increase variance
    - Extremely low variance → may increase bias
- So the sweet spot is usually **low bias + moderate variance**.

Low bias, high variance → Overfit (not ideal)

Low bias, low variance → **Best fit (ideal)**

High bias, high variance → Worst fit

High bias, low variance → Underfit

Correct: Low bias and low variance

ias-Variance Tradeoff

- High Bias: Model is too simple, can't capture underlying patterns (underfitting)
  - Example: Linear regression on non-linear data
- High Variance: Model is too complex, fits noise in training data (overfitting)
  - Example: High-degree polynomial regression on noisy data
- Low Bias, Low Variance: Model is just right, captures patterns and generalizes well
  - Example: Regularized regression (e.g., Lasso, Ridge) on suitable data

Consequences

- Underfitting: Poor performance on training and test data

- Overfitting: Good performance on training data, poor performance on test data
- Best Fit: Good performance on both training and test data

Strategies to Balance Bias and Variance

1. Regularization: Add penalty terms to reduce model complexity (e.g., L1, L2 regularization)
2. Cross-validation: Evaluate model on unseen data to detect overfitting
3. Early Stopping: Stop training when performance on validation set starts to degrade
4. Data Augmentation: Increase dataset size to reduce overfitting
5. Model Selection: Choose a model with suitable complexity for the problem