18-661: Introduction to ML for Engineers

Bias-Variance Tradeoff Review

Spring 2025

ECE – Carnegie Mellon University

In-class quiz

Given: Dataset $(x^{(i)}, y^{(i)})$ with regression line:

$$y = w_1 x + b_1 \tag{1}$$

Suppose we transform the dataset by shifting both inputs and outputs:

$$(x^{(i)} + \alpha, y^{(i)} + \beta) \tag{2}$$

where $\alpha > 0, \beta > 0, w_1\alpha \neq \beta$. The new regression line is $y = w_2 x + b_2$.
How does the transformation affect the regression equation?

- Original regression equation: $y^{(i)} = w_1 x^{(i)} + b_1$
- Transformed data: $y^{(i)} + \beta = w_2(x^{(i)} + \alpha) + b_2$

Expanding,

$$w_1 x^{(i)} + b_1 + \beta = w_2 x^{(i)} + w_2\alpha + b_2 \tag{3}$$

In-class quiz

Comparing coefficients of $x^{(i)}$ and constants,

$$
\begin{aligned}
w_1 = w_2 \Rightarrow w_2 &= w_1 &(4)\\
b_1 + \beta &= w_1\alpha + b_2 &(5)\\
\Rightarrow b_2 &= b_1 - w_1\alpha + \beta &(6)
\end{aligned}
$$

The correct answer is:

$$w_2 = w_1, \quad b_2 = b_1 - w_1\alpha + \beta \tag{7}$$

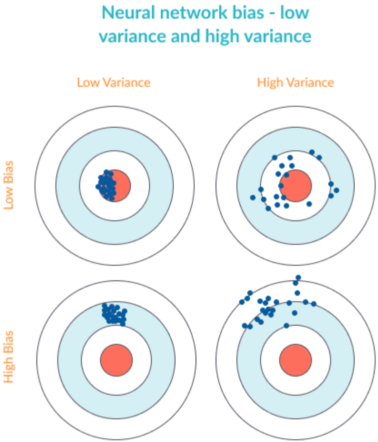
Bias-Variance Tradeoff

Definition: The bias-variance tradeoff refers to the balance between two sources of error in a model:

- **Bias:** Error introduced by approximating a real-world problem with a simplified model.
- **Variance:** Error introduced by the model's sensitivity to small fluctuations in the training set.

Key Insight: Increasing model complexity reduces bias but increases variance, and vice versa.

Bias and Variance Example



- **Overfitted:** Shoots at the target on average (low bias) but makes mistakes all the time (high variance).
- **Underfitted:** Shoots away from the target (high bias) but does so consistently (low variance).

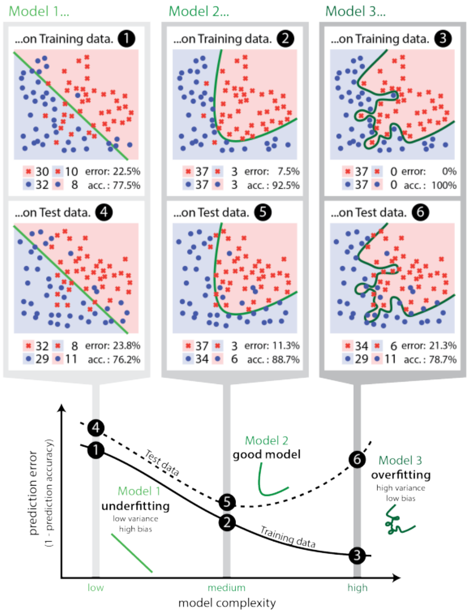
Remember this figure!

Example of Bias-Variance Tradeoff

Example: Polynomial Regression

- **High Bias:** A linear model (degree 1) may underfit the data, leading to high error.
- **High Variance:** A high-degree polynomial model may overfit, capturing noise instead of the underlying trend.
- **Optimal Tradeoff:** A moderate-degree polynomial balances bias and variance, achieving better generalization.

Example of Bias-Variance Tradeoff



Ridge Regression and the Bias-Variance Tradeoff

Ridge regression introduces an L_2 penalty to control model complexity:

$$\hat{w} = \arg\min_{w} \sum_{i=1}^{n} (y_i - w^T x_i)^2 + \lambda \|w\|^2 \tag{8}$$

- Increasing λ reduces model complexity, decreasing variance but increasing bias.
- Smaller λ allows more flexibility, reducing bias but increasing variance.

Proof: Ridge Regression Increases Bias and Decreases Variance

Let X be the feature matrix, y be the target, and $\hat{w}$ the estimated weights.

- The ordinary least squares (OLS) estimate is:

$$\hat{w}_{OLS} = (X^T X)^{-1} X^T y \tag{9}$$

- Ridge regression modifies this to:

$$\hat{w}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \tag{10}$$

Effect on Bias:

- Ridge shrinks weights toward zero, reducing model flexibility.
- This introduces bias as the model cannot perfectly fit training data.

Effect on Variance:

- Regularization stabilizes the inverse matrix, reducing sensitivity to fluctuations in training data.
- Lower variance improves generalization.

Proof: Ridge Regression Increases Bias and Decreases Variance in Predictions

Definition of Bias and Variance:

- **Bias:**

$$\text{Bias}^2 = \mathbb{E}[(\mathbb{E}[\hat{y}] - y)^2] \tag{11}$$

- **Variance:**

$$\text{Var}(\hat{y}) = \mathbb{E}[(\hat{y} - \mathbb{E}[\hat{y}])^2] \tag{12}$$

Effect on Bias:

- The ridge estimator is given by:

$$\hat{w}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \tag{13}$$

- The expectation of the ridge predictor is:

$$\mathbb{E}[\hat{y}_{ridge}] = X(X^T X + \lambda I)^{-1} X^T \mathbb{E}[y] \tag{14}$$

- Since the transformation shrinks coefficients, the expected prediction deviates from the true values, increasing bias.

Proof: Ridge Regression Increases Bias and Decreases Variance in Predictions

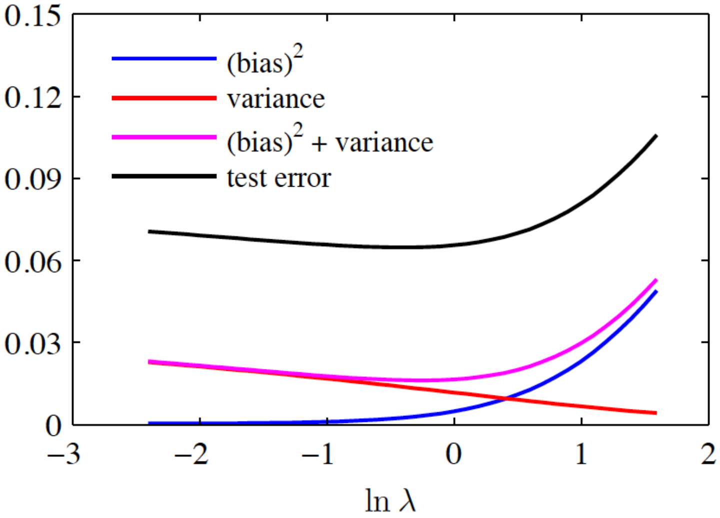
Definition of Bias and Variance:

- **Bias:**
$$\text{Bias}^2 = \mathbb{E}[(\mathbb{E}[\hat{y}] - y)^2] \tag{15}$$
- **Variance:**
$$\text{Var}(\hat{y}) = \mathbb{E}[(\hat{y} - \mathbb{E}[\hat{y}])^2] \tag{16}$$

Effect on Variance:

- The variance of ridge regression predictions is:
$$\text{Var}(\hat{y}_{ridge}) = X(X^TX + \lambda I)^{-1}X^T\sigma^2 IX(X^TX + \lambda I)^{-1}X^T \tag{17}$$
- Since $(X^TX + \lambda I)^{-1}$ stabilizes the inversion, the variance is reduced compared to ordinary least squares (OLS).

Remember this figure



Conclusion

Key Takeaways:

- Bias-variance tradeoff is crucial for model performance.
- Ridge regression increases bias but decreases variance in predictions.
- Choosing an appropriate regularization strength (λ) is essential for balancing bias and variance.