

MA477: Data Science
Lesson 1 Board Sheet — 13 January 2026
 United States Military Academy, West Point
 Instructor: MAJ Patrick Kuiper

1 Lesson Objectives

By the end of this lesson, cadets will be able to:

- Understand the steps in completing an end-to-end machine learning project.
- Understand how to implement data pipelines, clean/standardize data, visualize data, how to handle categorical/text
- Understand and know how to perform cross-validation

2 Statistical Learning Intuition

2.1 Model Selection

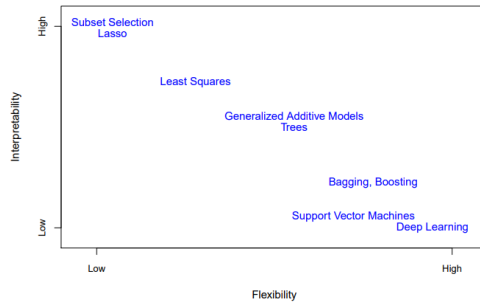


FIGURE 2.7. A representation of the tradeoff between flexibility and interpretability, using different statistical learning methods. In general, as the flexibility of a method increases, its interpretability decreases.

Figure 1

2.2 Bias Variance Considerations

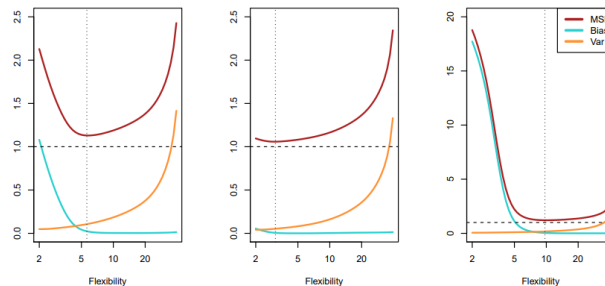


FIGURE 2.12. Squared bias (blue curve), variance (orange curve), $\text{Var}(\epsilon)$ (dashed line), and test MSE (red curve) for the three data sets in Figures 2.9–2.11. The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.

Figure 2

Error Metrics and Their Interpretation

1. Most Common Error Metric

What is the most commonly used error metric in regression problems, and why is it so widely adopted in practice?

2. Alternative Error Metric

What is another commonly used error metric, and in what situations might it be preferred over the most common choice?

3. Tradeoffs Between Metrics

What are the primary advantages and disadvantages of these two error metrics, particularly with respect to sensitivity to large errors and robustness to outliers?

4. Analytical Demonstration

How can the difference between these two error metrics be demonstrated analytically, for example by comparing their geometric interpretations or their level sets in two dimensions?