MA477: Data Science

Lesson 1 Board Sheet — 6 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 general model $Y = f(X) + \epsilon$.
- Understand how to measure model accuracy (loss function).
- Understand the difference between prediction and inference.
- Understand the difference between parametric and nonparametric models.
- Recognize situations when supervised vs unsupervised learning is appropriate.
- Explain the bias-variance trade-off.

2 Linear Algebra Warm-Up (By Hand)

Purpose: Linear algebra is the language of data science. Today we focus on intuition, structure, and meaning — not proofs or computation speed.

2.1 Matrix Dimensions

Given the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

- How many rows does A have?
- How many columns does A have?
- What is the dimension of A?

Check: What dimensions must another matrix have so that multiplication with A is valid?

2.2 Matrix Multiplication: 3×2 times 2×2

Let

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$$

- What are the dimensions of A and B?
- Is the product AB defined?
- What will the dimensions of AB be?
- Compute AB by hand.

Board Discussion: Each row of A is transformed by B. What does that suggest about matrix multiplication as a transformation?

2.3 Transpose Example

Given

$$C = \begin{bmatrix} 1 & 4 & -2 \\ 0 & 3 & 5 \end{bmatrix}$$

- What is the dimension of C?
- Compute C^T.
- What are the dimensions of C^T?

Interpretation:

- Rows become columns
- Transpose changes how we interpret inputs vs outputs
- This will matter later for dot products and projections

2.4 Solving for Variables Using Matrix Multiplication

Consider the system

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 5 \\ 3 \end{bmatrix}}_{\mathbf{b}}$$

- What does this matrix equation represent?
- Rewrite it as two scalar equations.

Now suppose we multiply both sides by the inverse of the coefficient matrix:

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$$

$$\mathbf{x} = A^{-1}\mathbf{b}$$

- Why does this isolate the unknown vector?
- What assumptions must be true for A^{-1} to exist?

Optional (By Hand): Compute A^{-1} and solve for (x, y):

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad A^{-1} = \frac{1}{(2)(1) - (1)(1)} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

2.5 Linear Independence: Dependent vs Independent

Example 1: Linearly Dependent Rows

$$D = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

- Is one row a scalar multiple of the other?
- What does this imply about the information content?

Example 2: Linearly Independent Rows

$$E = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$$

- Can either row be written as a multiple of the other?
- How many independent directions exist?

2.6 Why Linear Independence Matters (Analysis)

Board Analysis: Compare matrices D and E.

- Which matrix loses information? Explain in words.
- Which matrix can be inverted? Why?
- How does linear dependence affect our ability to solve systems $A\mathbf{x} = \mathbf{b}$?
- In a data setting: what might it mean if two features are linearly dependent?

Data Science Connection (Preview):

- Dependent features provide redundant information
- Independent features add explanatory power
- Later: dependence can cause unstable models and non-unique solutions

3 Statistical Learning Intuition

3.1 Model Selection

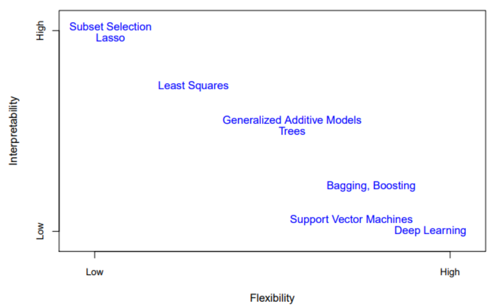

FIGURE 2.7. *A representation of the tradeoff between flexibility and interpretability, using different statistical learning methods. In general, as the flexibility of a method increases, its interpretability decreases.*

Figure 1

3.2 Bias Variance Considerations

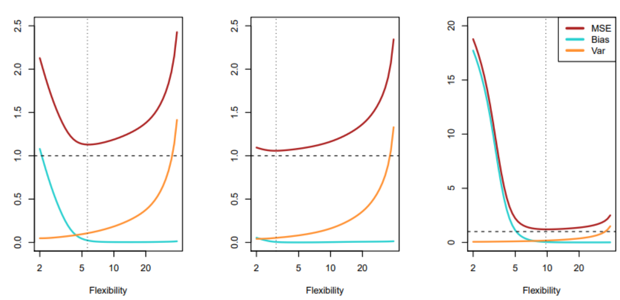

FIGURE 2.12. *Squared bias (blue curve), variance (orange curve), Var(ϵ) (dashed line), and test MSE (red curve) for the three data sets in Figures 2.9–2.11. The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.*

Figure 2