MA477: Data Science

Lesson Outline — 13 January 2026

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

1 Administrative

- Problem Set 1 due on 21 Jan
- Meeting two more times this week
- After review, review on sources of error, parametric models, and norms, on to notebooks

2 Student Review

3 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

4 Parametric Model Review

Table 1: Conceptual Comparison of Parametric and Non-Parametric Models

Characteristic	**Parametric Models**	**Non-Parametric Models**
Model form	Fixed functional form	No fixed functional form
Number of parameters	Fixed; does not grow with data	Grows with data or model complexity
Assumptions	Strong assumptions about structure or distribution	Minimal assumptions about data
Flexibility	Limited	High
Data requirements	Lower	Higher
Bias–variance tendency	Higher bias, lower variance	Lower bias, higher variance
Interpretability	Often high	Varies; often lower
Computational cost	Typically low	Often higher
Typical modeling question	"Does a simple relationship explain the data?"	"What structure does the data suggest?"

Table 2: Examples of Parametric and Non-Parametric Models

Model	**Category**	**Reason for Classification**
Linear Regression	Parametric	Assumes a linear relationship with a fixed number of coefficients
Logistic Regression	Parametric	Uses a fixed sigmoid function with a fixed parameter set
Linear Discriminant Analysis (LDA)	Parametric	Assumes class-conditional normal distributions
Naive Bayes	Parametric	Assumes specific probability distributions and conditional independence
k-Nearest Neighbors (k-NN)	Non-Parametric	No explicit model; predictions depend directly on stored data
Decision Trees	Non-Parametric	Tree structure adapts to data with no fixed size
Random Forests	Non-Parametric	Ensemble of adaptive decision trees
Kernel Density Estimation (KDE)	Non-Parametric	Estimates density directly from data without distributional assumptions
Gaussian Processes	Non-Parametric	Places a distribution over functions rather than parameters
LOESS / LOWESS	Non-Parametric	Fits local models whose form depends on nearby data
Neural Networks	Semi-Parametric	Fixed architecture but expressive power grows with data and training

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?
   The most commonly used error metric is mean squared error (MSE). It is widely adopted because it is mathematically convenient, differentiable everywhere, and leads to closed-form solutions or efficient gradient-based optimization methods.

2. **Alternative Error Metric**
   What is another commonly used error metric, and in what situations might it be preferred over the most common choice?
   Mean absolute error (MAE) is another common error metric. It is often preferred when robustness to outliers is important, since it penalizes errors linearly rather than quadratically.

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?
   MSE strongly penalizes large errors, making it sensitive to outliers but effective when large deviations are especially costly. MAE treats all errors proportionally, making it more robust to outliers but less sensitive to rare large errors.

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?
   Analytically, MAE corresponds to the L1 norm and produces diamond-shaped level sets in two dimensions, while the square root of MSE corresponds to the L2 norm and produces circular level sets. These different geometries explain their distinct optimization and robustness properties.

For a vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the most common norms are:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \quad \text{(L1 norm: sum of absolute values)},$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \quad \text{(L2 norm: Euclidean length)},$$

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i| \quad \text{(L-infinity norm: maximum component)}.$$