Week 03: ML Models Fundamentals

From Smarter Data to Smarter Models

What Week 3 Covers

This week, we transition from data fundamentals to core modeling techniques for real-world ML. You'll learn to build, train, and evaluate your first predictive models.

01

Module 09: Advanced Preprocessing

Master outlier detection, polynomial transformations, and domain-driven feature engineering

02

Module 10: Regression Models

Learn linear and logistic regression, cost functions, and gradient descent optimization

03

Module 11: Decision Trees

Explore tree-based models, entropy, information gain, and pruning strategies

04

Practice Modules

Apply your knowledge with hands-on exercises in modules 10.5 and 11.5

05

End-of-Week Quiz

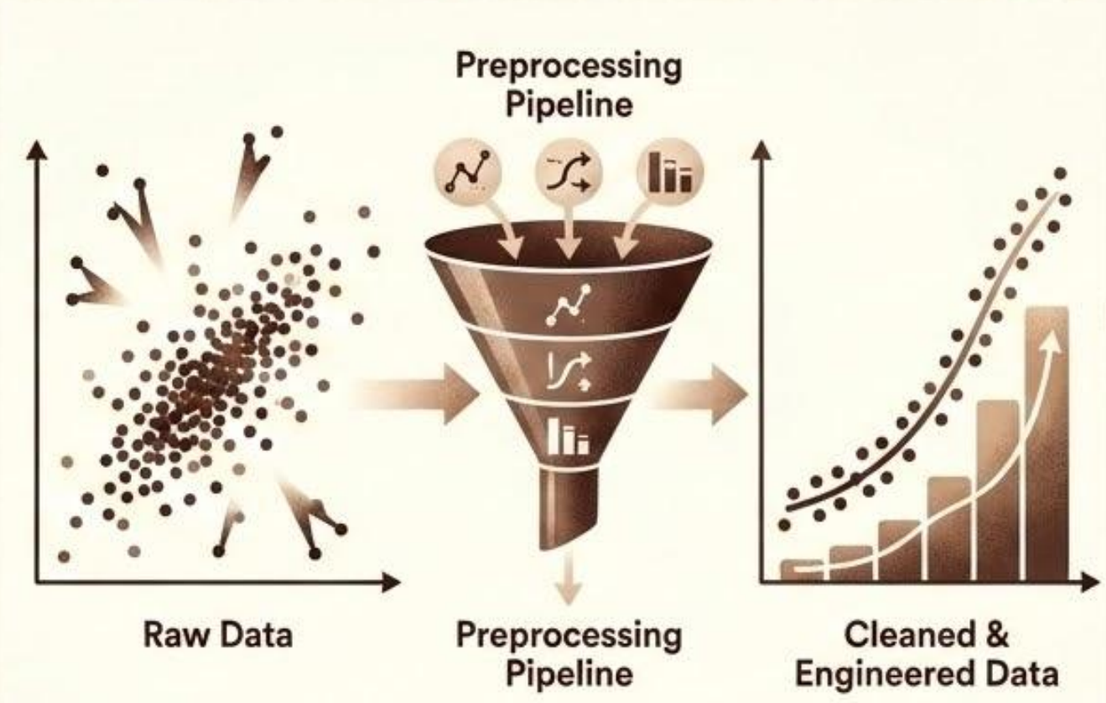
Test your understanding and solidify fundamental concepts before advancing

Module 09: Advanced Preprocessing & Feature Engineering

What You'll Learn

This module expands on Part 1, covering advanced data preparation techniques to significantly improve model performance.

Outlier Detection & Handling: Identify anomalies and decide whether to remove, cap, or transform them.
Polynomial Transformations: Create non-linear features to capture complex relationships in your data.
Binning Strategies: Convert continuous variables into categorical buckets for better interpretability.
Domain-Driven Features: Leverage subject matter expertise to engineer meaningful predictive variables.



Module 10: Linear Regression Fundamentals

Linear regression is the foundation of predictive modeling. Understanding how it works gives you intuition for nearly every ML algorithm that follows.

Linear Regression Basics

Learn how to fit a line to data points, predicting continuous outcomes from input features.

Cost Function Concept

Understand how models measure error using Mean Squared Error (MSE) and other loss functions.

Gradient Descent

Grasp the optimization algorithm that iteratively adjusts parameters to minimize prediction error.

Real-world example: Predict house prices using features like square footage, bedrooms, and location. Linear regression establishes the mathematical relationship between these inputs and price, enabling accurate predictions.

Logistic Regression: From Lines to Classifications

Logistic regression adapts the regression framework for classification, outputting probabilities (0-1) for binary decisions.

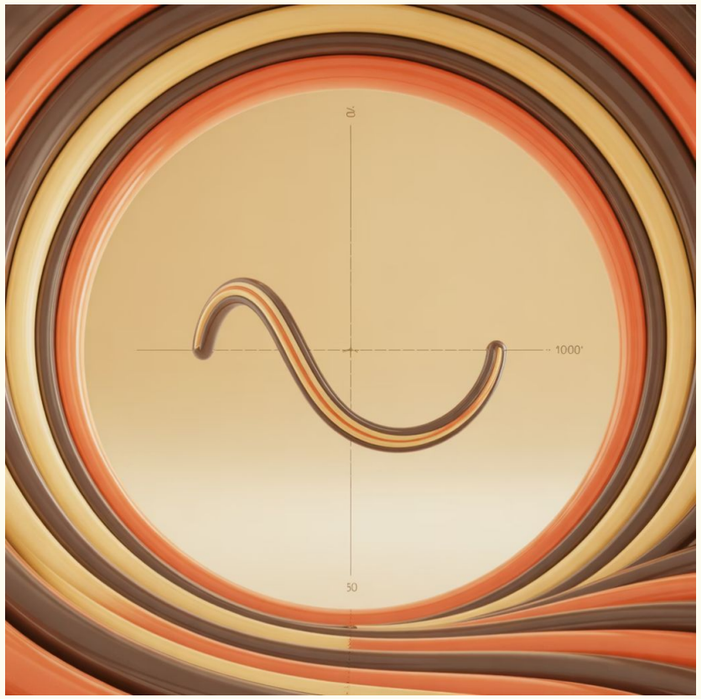
Key Concepts

Sigmoid Function: Transforms linear outputs into probabilities (0-1)

Decision Boundary: Threshold separating classes in feature space

Log Loss: Cost function optimized during training for accuracy

Common applications include: email spam detection, customer churn prediction, and medical diagnosis. It is ideal for any yes/no decision scenario.



Evaluating Classification Models

Building a classifier is only half the battle. Understanding its performance requires the right metrics, as different evaluation methods reveal different aspects of model quality.

Confusion Matrix

A table showing true positives, false positives, true negatives, and false negatives: the foundation of all classification metrics.

Precision & Recall

Precision measures the correctness of positive predictions; recall measures how many actual positives were caught.

F1 Score

The harmonic mean of precision and recall, providing a single balanced metric when both matter equally.

ROC Curve & AUC

The ROC Curve plots true positive rate vs. false positive rate across thresholds; AUC summarizes overall discrimination ability.

Module 10.5: Practice Module 9 & 10

Theory becomes expertise through practice. This module provides structured exercises to apply concepts from both Module 9 (Advanced Preprocessing) and Module 10 (Regression Models) to real-world datasets.

1

Real Datasets

Work with authentic data from domains like healthcare, finance, and e-commerce, not just toy examples.

2

End-to-End Workflows

Practice the complete ML pipeline: data loading, preprocessing, feature engineering, model training, and evaluation.

3

Troubleshooting

Diagnose common issues such as poor convergence, overfitting, and feature scaling problems.

Key Emphasis: This module merges the advanced preprocessing techniques from **Module 9 (Advanced Preprocessing)** with the regression models covered in **Module 10 (Regression Models)**. Discover how proper data preparation directly impacts model performance. Experiment with various feature transformations and compare results, applying the models you've learned.

Module 11: Decision Trees Deep Dive

Decision trees represent a fundamentally different approach to machine learning: they create hierarchical rules rather than mathematical equations. Their intuitive structure makes them both powerful and interpretable.

Tree Structure

Learn how trees split data recursively based on feature values, creating a flowchart-like decision path.

Splitting Criteria

Understand entropy, Gini impurity, and information gain: the metrics that determine optimal splits.

Pruning Strategies

Discover how to prevent overfitting by limiting tree depth and removing unnecessary branches.

Why Decision Trees Matter: Unlike linear models that assume smooth relationships, trees can capture complex, non-linear patterns with sharp boundaries. They're the foundation of powerful ensemble methods like Random Forests and Gradient Boosting.

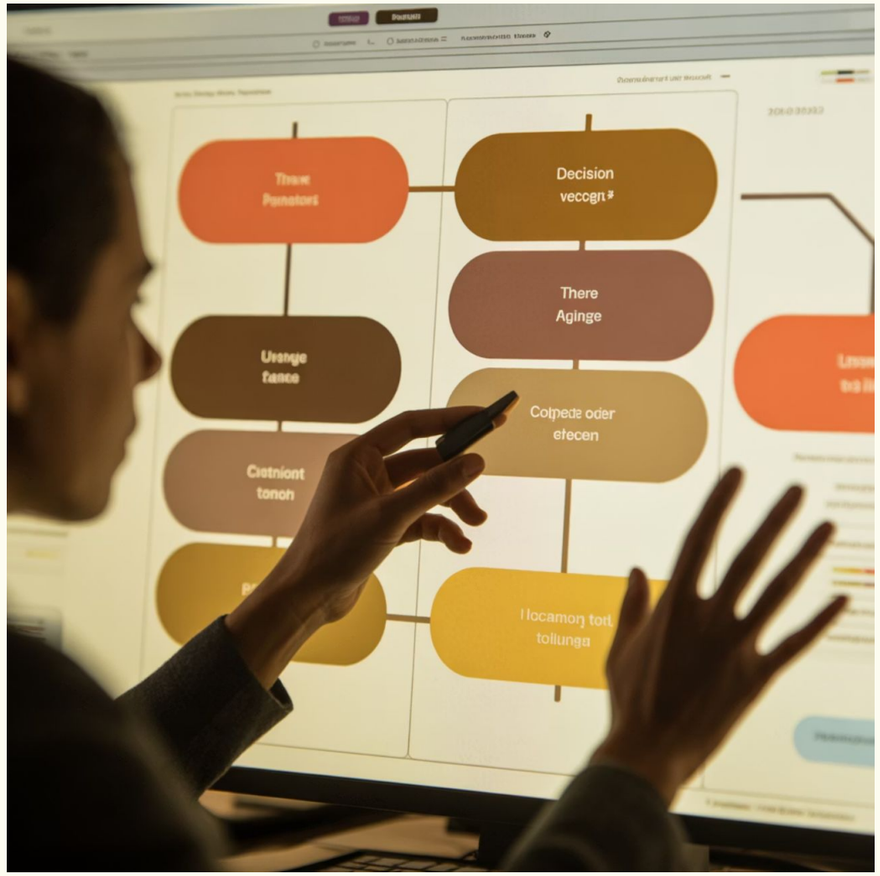
Module 11.5: Build Your Own Decision Tree

In this practice module, you'll construct decision trees from scratch and learn to interpret their predictions. The goal isn't just accuracy; it's understanding why the tree makes certain decisions.

Practice Activities

1. Build a decision tree classifier for customer segmentation
2. Visualize the tree structure and trace prediction paths
3. Experiment with different hyperparameters (max depth, min samples split)
4. Compare tree performance before and after pruning
5. Extract and interpret the most important features

Interpretation Challenge: Given a trained tree, explain its decisions to a non-technical stakeholder. Can you translate "Gini = 0.42" into business insight?



Module 12: End-of-Week Quiz

This quiz tests your understanding of all foundational concepts covered in Week 3.

You're Now a Model Builder

From Data to Decisions

What You've Mastered

Advanced preprocessing, linear and logistic regression, decision trees, and comprehensive model evaluation.

What's Next

Ensemble methods, neural networks, and deep learning, building on these fundamentals for more powerful models.

Call to Action: Don't skip the practice problems. Applying concepts comes from repetition and experimentation. Lock in these fundamentals now, and everything that follows will click into place.