Module 4: Machine Learning – Regression

Data Science with Python – BSc Course

Why Regression?

The Capital Asset Pricing Model (CAPM) is a regression. Fama-French's three-factor model, which reshaped how we understand equity returns, is a regression. House price indices, inflation forecasts, credit scoring models – regression underpins trillions of dollars in asset pricing and risk management.
Regression is not just a statistical technique. It is the foundation of quantitative finance, from pricing derivatives to forecasting GDP.

Regression models power the quantitative foundations of modern finance

Why This Matters

- **Asset pricing:** CAPM and factor models are built on linear regression
- **Credit risk:** Probability of default (PD) models use regression
- **Real estate:** Hedonic pricing models value properties with regression
- **Forecasting:** Interest rates, GDP, inflation all use regression models

Regression is the workhorse of quantitative finance

What You'll Learn

By the end of this module, you will be able to:

- Build and interpret linear regression models using scikit-learn
- Apply regularization (Ridge, Lasso, ElasticNet) to prevent overfitting
- Evaluate models with appropriate metrics (MSE, MAE, R-squared)
- Understand and implement factor models for asset pricing

From theory to production-ready regression models

Lesson Roadmap: Four Lectures, One Story

Lesson	Narrative Arc	Core Concept
L21	From Galton to Stock Betas	OLS, coefficients, CAPM
L22	The Goldilocks Problem	Ridge, Lasso, bias-variance
L23	Trust But Verify	MSE, MAE, R-squared, walk-forward
L24	One Factor Is Never Enough	Fama-French, factor attribution

The arc: One variable → many variables → measuring quality → multi-factor finance models. Each lecture builds on the previous. By L24, you combine everything into production factor models.

Each lesson: 45 minutes lecture + 25-minute hands-on exercise

Key Concepts

- **Linear Regression (OLS)** – Fit lines to data, interpret coefficients
- **Regularization (Ridge, Lasso, ElasticNet)** – Prevent overfitting with penalties
- **Model Evaluation (MSE, MAE, R-squared)** – Choose the right metric for your task
- **Factor Models (CAPM, Fama-French)** – Understand asset pricing through factors

Master regression and you have the foundation for all of machine learning

Finance Application

Scenario: Factor-Based Stock Screener
Using the skills from this module, you can:

- Build a Fama-French three-factor model to explain stock returns
- Identify stocks with positive alpha (excess returns)
- Use regularization to handle multicollinearity in factor models
- Evaluate model performance with cross-validation

This is the foundation of quantitative equity strategies used by AQR and Dimensional Fund Advisors.

Factor models are how quants separate skill from luck

Who Uses This?

- **Asset Management** – AQR, Dimensional Fund Advisors build factor portfolios
- **Banking** – Credit scoring uses logistic and linear regression
- **Real Estate** – Zillow's Zestimate is a regression model
- **Central Banks** – Fed's economic forecasts use regression models

Regression is everywhere in finance

What's Next: Module 5 – ML: Classification

Regression predicts continuous values (stock prices, interest rates). Classification predicts categories: fraud/not-fraud, default/no-default, buy/hold/sell.

Module 5 introduces logistic regression, decision trees, and classification metrics – the tools for categorical prediction in finance.

Prerequisite check: Can you build a regression model, apply regularization, and evaluate with cross-validation? If yes, you are ready.

From continuous predictions to categorical decisions

Let's Begin!

First up: L21 – "From Galton's Sweet Peas to Stock Betas"

The 150-year story of the best-fit line.