Lesson 21: Linear Regression

Data Science with Python – BSc Course

45 Minutes

Learning Objectives

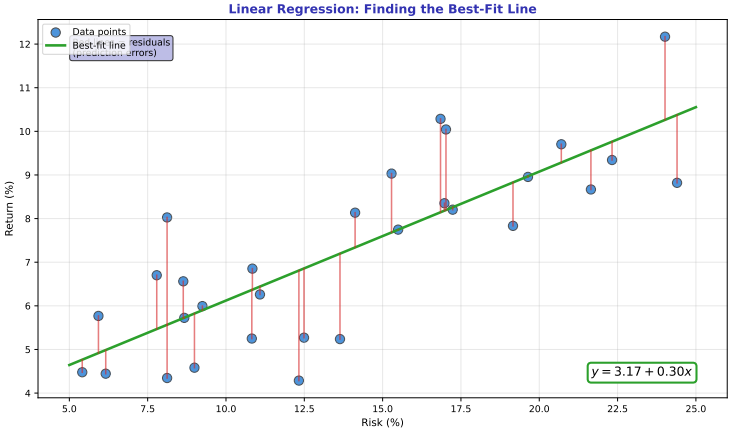
The Problem: How do we quantify systematic risk for portfolio construction?

After this lesson, you will be able to:

- Understand OLS estimation and the least squares principle
- Fit linear models using sklearn's LinearRegression
- Interpret slope (beta) and intercept (alpha)

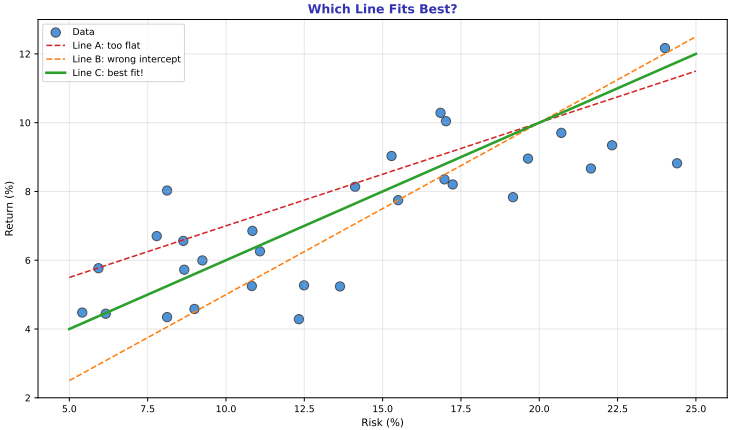
Finance Application: Stock classification by CAPM beta

Regression Concept



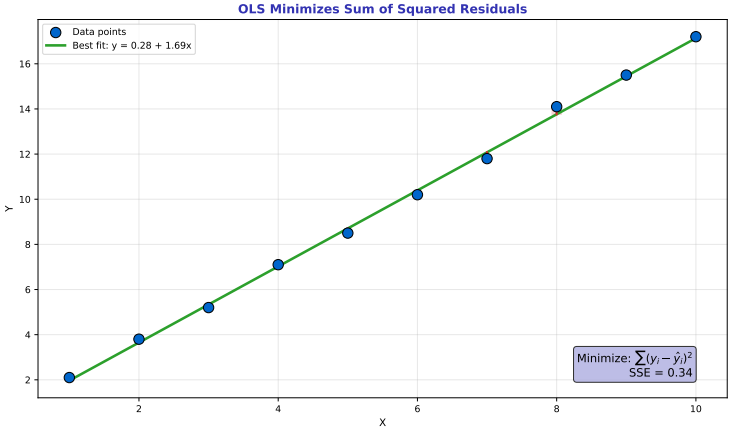
The best line minimizes the sum of squared vertical distances (residuals)

Which Line Fits Best?



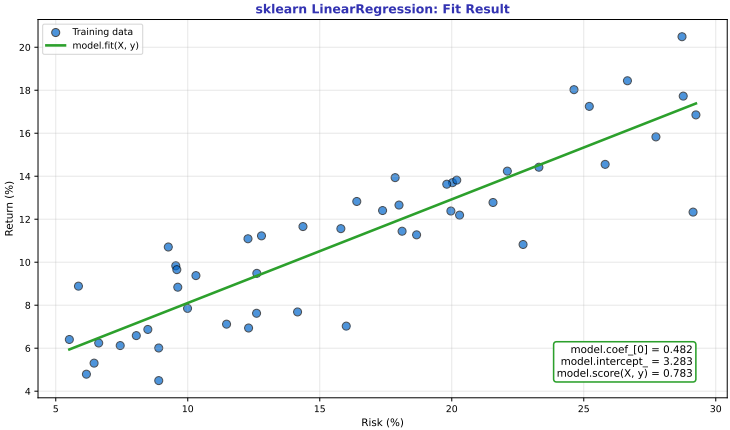
OLS finds the unique line that minimizes total squared error

OLS Formula



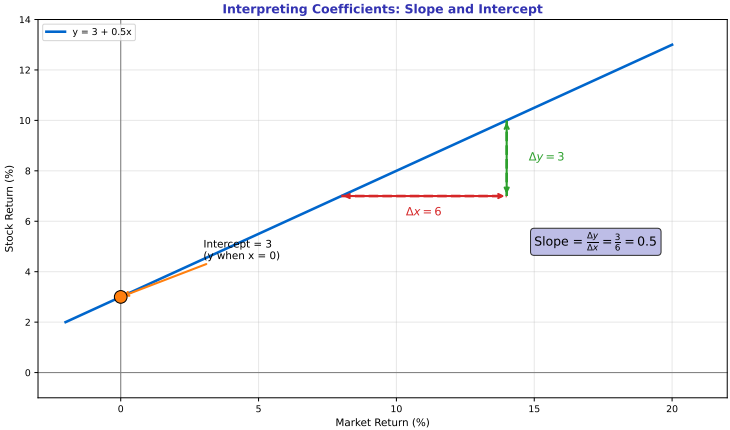
Why squares? (1) Makes all errors positive, (2) Penalizes large errors more

sklearn API



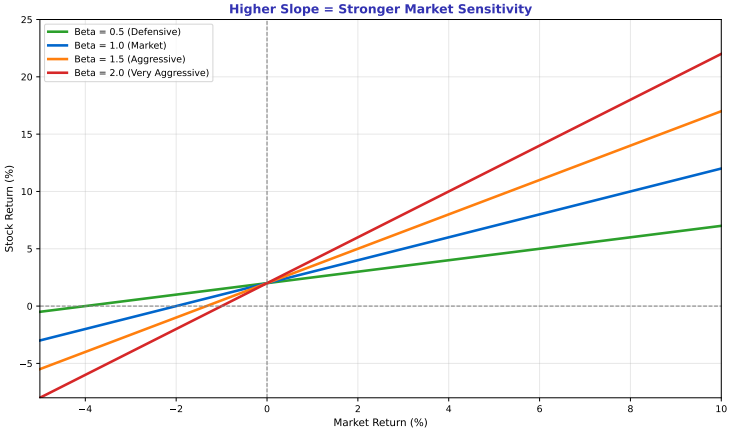
Pattern: model.fit(X, y) then model.predict(X_new) – works for all sklearn models

Coefficient Interpretation



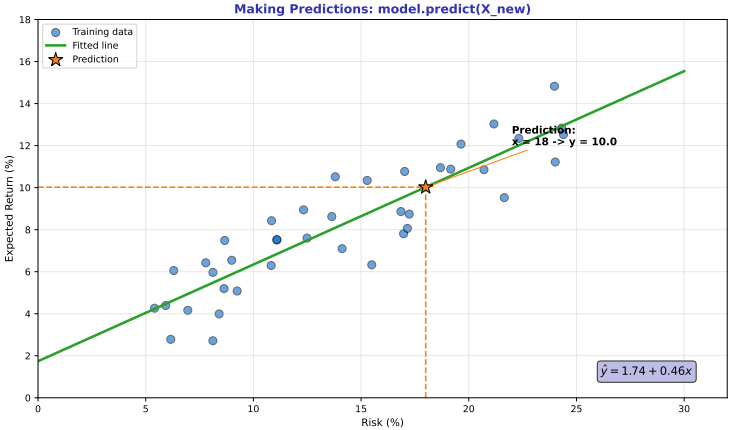
Slope = sensitivity to X. Intercept = value when X = 0

Different Slopes = Different Betas



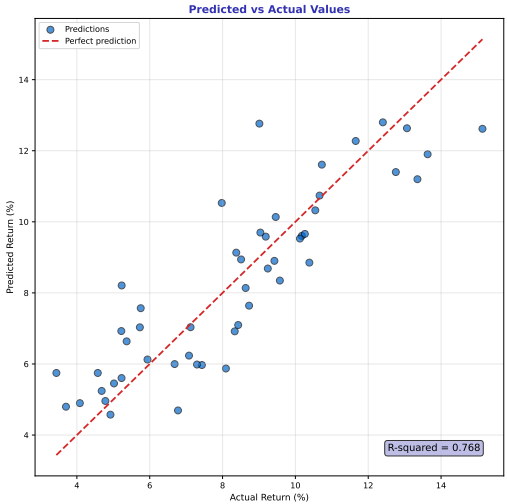
Higher beta = stock amplifies market moves more

Making Predictions



Prediction: plug X into the equation to get estimated Y

Predicted vs Actual

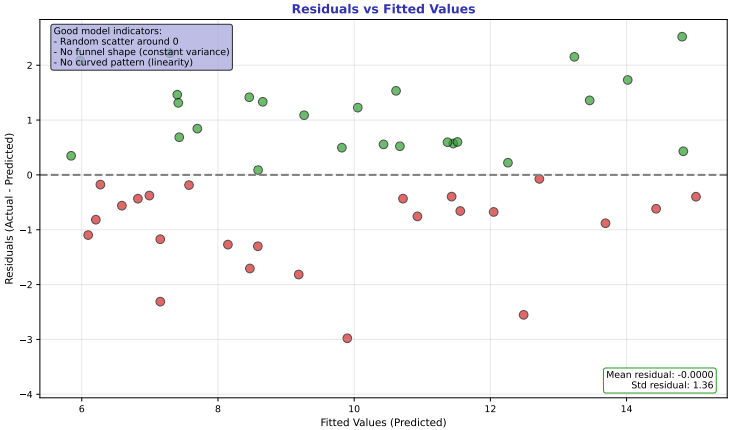


Points on the diagonal = perfect predictions

Extrapolation Warning

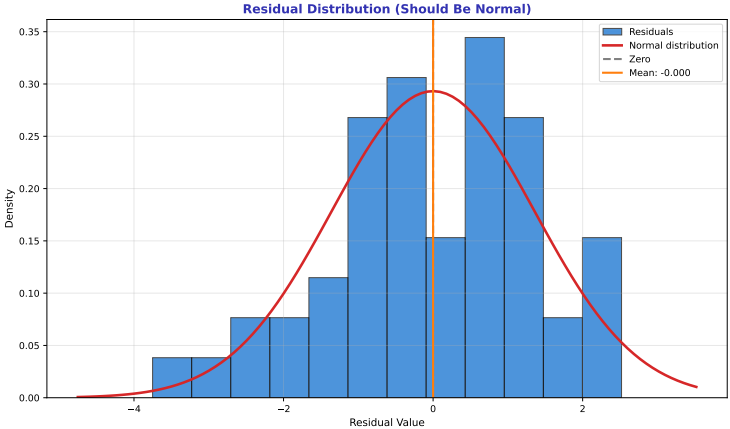
Never predict outside your training data range!

Residuals vs Fitted



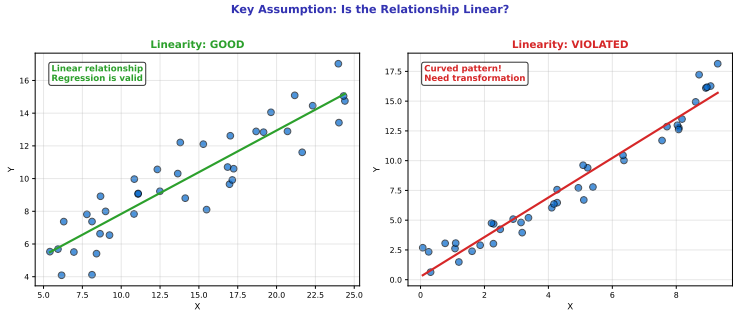
Random scatter = good model. Pattern = problem

Residual Distribution



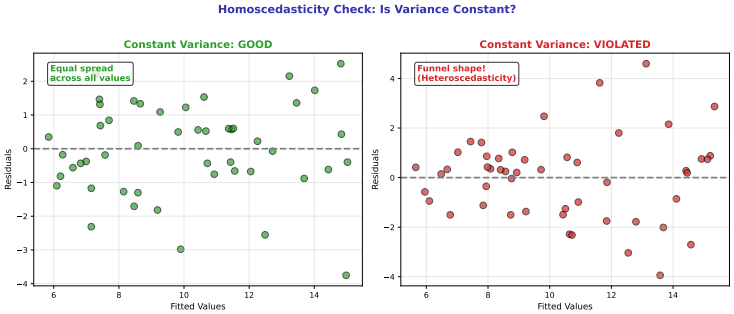
Normality assumption: residuals should follow a bell curve

Linearity Assumption



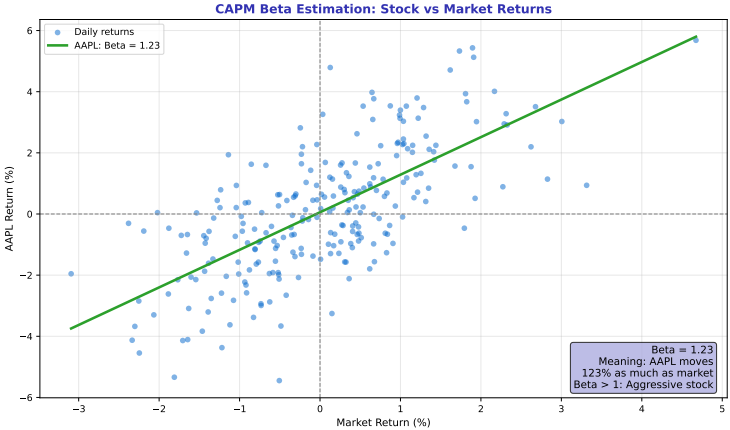
If curved pattern, try polynomial features or transformation

Homoscedasticity



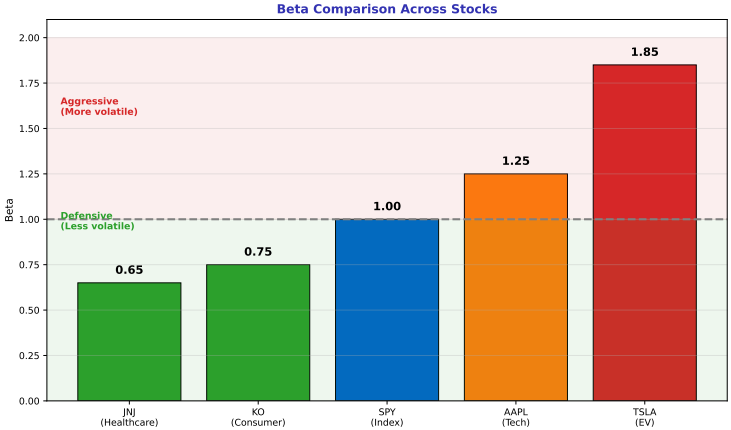
Funnel shape = heteroscedasticity. Fix: weighted least squares or log transform

CAPM Beta Estimation



Beta = slope of stock returns vs market returns

Beta Comparison



Mix defensive (low beta) and aggressive (high beta) stocks based on risk tolerance

Hands-On Exercise (25 min)

Task: Estimate Beta for Your Favorite Stock

1. Download 1 year of daily returns for a stock and SPY
2. Fit: `model.fit(spy_returns, stock_returns)`
3. Extract beta and alpha from coefficients
4. Plot the regression line with data points

Deliverable: Scatter plot with regression line, annotated with beta value.

Extension: Compare 1-year vs 5-year beta estimates

Lesson Summary

Problem Solved: Quantify systematic risk via CAPM beta

Key Takeaways:

- OLS minimizes squared errors
- sklearn: `LinearRegression().fit(X, y)`
- Slope = beta, Intercept = alpha

Next: Regularization (L22) – handling too many features

Memory: Beta = slope. High beta = high volatility.