Week 9 — Machine Learning for Forecasting

Regularisation, Trees, and the Bias-Variance Tradeoff

Dr Juan Zurita

Economic & Business Forecasting
The University of Edinburgh · School of Economics

The Bias-Variance Tradeoff

Why ML for Forecasting?

- Traditional models: low bias, potentially high variance when p is large
- ML insight: accept some bias to reduce variance $\Rightarrow$ lower total error
- Especially useful with **many predictors** relative to sample size

$$\text{MSE} = \underbrace{\text{Bias}^2}_{\text{model too simple}} + \underbrace{\text{Variance}}_{\text{model too complex}} + \underbrace{\sigma_\varepsilon^2}_{\text{irreducible}}$$

Regularised Regression

Ridge and Lasso

Ridge (L_2 penalty):

$$\hat{\beta}_{\text{Ridge}} = \arg\min \|y - X\beta\|^2 + \lambda\|\beta\|_2^2$$

Shrinks all coefficients towards zero.

Lasso (L_1 penalty):

$$\hat{\beta}_{\text{Lasso}} = \arg\min \|y - X\beta\|^2 + \lambda\|\beta\|_1$$

Sets some coefficients exactly to zero — automatic variable selection!

Elastic Net: combines both penalties.

Cross-Validation for λ

1. Split training data into K folds
2. For each λ, compute average out-of-fold MSE
3. Select λ with lowest CV error

Caution for time series: use **rolling-window** or **expanding-window** CV to respect temporal order.

Tree-Based Methods

Decision Trees and Random Forests

Decision tree:

- Recursively partition the feature space
- Predict the mean of each partition
- Nonlinear and interpretable, but high variance

Random forest:

- Average over B trees, each on a bootstrap sample
- Each split considers a random subset of features
- Low variance, retains nonlinearity
- Feature importance: how much does each lag matter?

Summary

- **Bias-variance tradeoff** is the key ML insight
- **Ridge**: shrinks; **Lasso**: shrinks + selects
- **Random forests**: nonlinear, robust
- Use time-aware cross-validation
- Next week: neural networks and deep learning