Kiara Alejandro

Professor Brasic

ECON 418

December 15, 2024

Final Exam

1. (i) Outputs (like housing prices) are predicted via supervised learning using labelled data. Unsupervised learning like customer clustering finds trends in unlabelled data. Unsupervised learning reveals data structure, while supervised learning predicts.

   (ii) Classes predict discrete outcomes like spam detection, while regression predicts continuous outcomes like housing prices. Logistic or Decision Trees classify, Linear Regression regresses.

   (iii) Non-reducible error is data noise. The cross-validation-estimated reducible error derives from model defects. k-Nearest Neighbours evaluates model performance on unseen data to predict test error.

   (iv) Unlike non-parametric methods like Decision Trees, parametric methods like Linear Regression presume a functional form. Compared to parametric models, non-parametric models are flexible yet computationally costly.

   (v) The training phase builds the model using labeled data, whereas the testing phase tests it on unseen data. Only after training is the testing set used. Tuning with the testing set risks overfitting; use a validation set.

   (vi) The majority vote combines model forecasts to decide. Each tree votes and the most common forecast is generated in random forests, reducing bias and volatility.

   (vii) It balances underfitting (high bias) with overfitting (high variance). It optimizes model complexity. Machine learning aims for models to generalize successfully on unseen data.

   (viii) Split data into 5 equal folds. Once each fold is tested, the other four are trained. All folds are averaged for robust error estimates and overfitting risk reduction.

   (xi) Decision trees use feature thresholds to subset data. Multiple decision trees trained on bootstrapped data and averaged forecasts reduce overfitting in random forests. Random forests infer causality using feature importance or variable significance.

Kiara Alejandro

Professor Brasic

ECON 418

December 15, 2024

2. (i)

$$L(\beta) = \|y - X\beta\|^2 + \lambda\|\beta\|^2$$

$$\rightarrow \|y - X\beta\|^2 \rightarrow (y - X\beta)'(y - X\beta)$$

$$\rightarrow L(\beta) = (y - X\beta)'(y - X\beta) + \lambda\beta'\beta$$

$$\frac{\partial L}{\partial \beta} = -2X'y + 2X'X\beta + 2\lambda\beta = 0$$

$$\rightarrow X'X\beta + \lambda I_K\beta = X'y$$

$$\rightarrow \beta = (X'X + \lambda I_K)^{-1}X'y$$

$$\rightarrow \underline{\underline{\hat{\beta}_R = (X'X + \lambda I_K)^{-1}X'y}}$$

(ii) Ridge regression leads to shrinkage bias. As λ grows, coefficient estimates $\hat{\beta}_R$ approach zero. The bias occurs when the ridge penalty λI_K decreases coefficients in the OLS solution to reduce variance. When prioritising predictive accuracy over unbiased estimates, machine learning accepts this bias since it avoids overfitting and increases generalization on unseen data. In econometrics, causal inference requires unbiased estimates, hence bias is problematic. In high-dimensional models or those with multicollinearity, ridge regression can stabilize estimates.

3. (i) The appropriate model is the Difference-in-Differences (DiD) regression:

$$\beta_0 + \beta_1 NJ_i + \beta_2 Post_t + \beta_3 (NJ_i \cdot Post_t) + u_{it}$$

NJ represents New Jersey (treatment group), Post represents November post-policy, and the interaction term represents the minimum wage increase's causal effect. Because the NJ variable accounts for the average difference between New Jersey and Pennsylvania, state-fixed effects are not included.

Kiara Alejandro

Professor Brasic

ECON 418

December 15, 2024

(ii) The table will show the average total employment for each state (0 = Pennsylvania, 1 = New Jersey) in February and November before and after the policy. We can track treatment and control group employment changes using these averages.

(iii) Subtract Pennsylvania's difference from New Jersey's between February and November to estimate the DiD using sample means. Treatment effect (ATT). DiD evaluates the minimum wage increase's causal effect on employment. The outcome is the "Average Treatment Effect on the Treated (ATT)."

(iv) When the CI does not include zero, reject the null hypothesis that ATT = 0. Check the CI to reject the null hypothesis if ATT = 5.

(v) To test the DiD estimator's parallel trends assumption, compare New Jersey and Pennsylvania employment trends before February 1992. Estimate a regression with an interaction term (state * time_period_before). Statistically insignificant interaction terms support the assertion. Visualizing pre-treatment employment trends for both states can also be confirmed.

(vi) Violation of the parallel trends assumption by this confounding factor could alter the DiD estimation. The announcement could depress fast-food demand and employment irrespective of the minimum wage change, bypassing the estimate and misattributing demand-driven effects to policy.

(vii) Time-invariant restaurant differences are accounted for by fixed effects. The DiD estimate may alter when these fixed effects compensate for restaurant-specific heterogeneity. Thus, the calculated treatment impact focuses on within-restaurant changes over time, potentially minimizing bias.

(viii) If DiD assumptions, especially parallel trends, hold, the estimate is trustworthy. The estimate may be biased if unobserved shocks influence employment separately. To improve reliability, we could evaluate pre-treatment trends or include robustness tests like alternate control groups or covariates.

(ix) Using propensity score matching, the effect can be estimated non-parametrically. This technique estimates treatment effects by comparing treated and control units with

Kiara Alejandro

Professor Brasic

ECON 418

December 15, 2024

similar features. The model definition assumptions are reduced and effects from comparable groups are directly identified.

(x) https://github.com/kiaraalejandro/kiaraalejandro.git