Notes on Papers

Trey Livian

October 2024

1 "Measuring the Completeness of Economic Models" by Fudenberg, et al

- Completeness := the fraction of the predictable variation in the data that the model captures.
- Loan Demand Example
  - If the best predictions are not much better than those of the NPV model, then alternative models built on the same features cannot possibly do much better on this data set. For new models to help, they must identify new variables that are not currently measured.
- Model Prediction Error Decomposition
  - Intrinsic noise in the outcome due to limitations of the features we have measured, that is, the irreducible error. This is an upper bound for the performance of the model based upon the features provided.
  - Regularities in the data that the model misses.
- Measuring Completeness $\frac{\epsilon_{\text{base}} - \epsilon_{\text{model}}}{\epsilon_{\text{base}} - \epsilon_{\text{irreducible}}}$
  - ϵ_{base} is the out of sample prediction error
  - ϵ_{model} is the out of sample error of the model
  - ϵ_{base} is the irreducible error
  - Baseline model could be something like guessing the outcome at random
  - Completeness is the model's reduction in prediction error relative to the baseline, divided by the achievable reduction in prediction error.
  - Interpretation of the Extreme Values of Completeness:
    - O means there is no improvement in the model over the baseline model
    - 1 means the model eliminates all reducible error.

- Important Note: A model can be complete for the given set of features even if it predicts poorly and its R^2 is low.
- Main challenge is measuring irreducible error.

- Measuring Irreducible Error
  - The performance of "black-box" machine learning methods can be used as a stand-in
  - You can obtain this by doing a "look-up" table, which is a non-parametric model selection (and possibly parameter tuning procedure) that yields the model with the highest out-of-sample predictive accuracy for those selected features.

- Applications to Illustrate Important Points
  - The absolute error of a model can be misleading.
  - A model's gain over the baseline model can be misleading.
  - The completeness measure reveals the relative complexity of various prediction problems.

- Completeness in the Context of Prediction Problems
  - The theoretical ideal prediction rule that minimizes the loss function will have an expected error that is equal to the irreducible error.
  - We call the ratio between the reduction in prediction error achieved by choosing the model divided by the achievable reduction in prediction error "Completeness."
  - Completeness Equation in Context: $\frac{E_P(f_{\text{base}})-E_P(f^*_\Theta)}{E_P(f_{\text{base}})-E_P(f^*)}$
    * $E_P(f_{\text{base}})$ is the expected prediction error of the baseline model.
    * $E_P(f^*_\Theta)$ is the expected prediction error of the best-performing element of the model class F^*_Θ.
    * $E_P(f^*)$ is the expected prediction error of the theoretically best model (equivalent to the irreducible error)

- Choice of Baseline Model
  - Sometimes there is a natural choice for the baseline model.
  - Completeness is stable across different choices.
  - If one wanted to take a less parametric approach one could opt to use the best unconditional prediction of Y. This depends on the loss function. In the case of the loss function being MSE, the average would be appropriate. The advantage of this would be that it eliminates arbitrary choices between different studies.
  - Disadvantages of non-parametric approach:

        * May not be in the same class of models as the model in question.
        * Very sensitive to variability in the data.

- Completeness is dependent on Feature Space (can be a proxy for Complexity of data). This can be seen as a strength of the approach because it provides nuance that certain types of models are better for certain data sets.

- Expanding the feature space is one of the only ways to improve a nearly complete model's predictive performance.

- Does not help to inform trade-offs in model performance with these metrics:

    – Parsimony: minimizing number of parameters.
    – Portability: performance in different settings.
    – Causal Explanations

- In general: the more flexible a model, the more complete it is.

- Completeness can help us see how much predictive power we gain for using a black box model versus a model that is easier to interpret.

- Necessary Assumption: There are a large number of observations.

- Estimator for Expected Prediction Errors:

    – Split Data into 10 disjoint Folds. Where individual folds are testing data and pairwise unions of disjoint data are training data.
    – Select prediction rule f that best fits the data. The in-sample performance of f for predicting observations in the training data is given by (sample analog of expected prediction error that was previously discussed): $e(f, Z^i_{\text{train}}) = \frac{1}{|Z^i_{\text{train}}|} \sum_{(x,y)\in Z^i_{\text{train}}} \ell(f(x), y)$
    – Choose the f_{i} that minimizes the expected error.
    – Evaluate how it performs out of sample: $CV_i = e(f_i, Z^i_{\text{test}})$
    – Average over out-of-sample errors (across K test sets): $CV(F, \{Z^i\}^{10}_{i=1}) = \frac{1}{K} \sum_{i=1}^{K} CV_{\text{i}}$

- Estimator for Completeness:

    – Define:

$$\hat{\epsilon}_{\text{base}} \equiv CV(f_{\text{base}}, \{Z^i\}^{10}_{i=1})$$
$$\hat{\epsilon}_{\Theta} \equiv CV(F_{\Theta}, \{Z^i\}^{10}_{i=1})$$
$$\hat{\epsilon}_{\text{best}} \equiv CV(X^Y, \{Z^i\}^{10}_{i=1})$$

– Compute:

$$\frac{\hat{\epsilon}_{\text{base}} - \hat{\epsilon}_{\Theta}}{\hat{\epsilon}_{\text{base}} - \hat{\epsilon}_{\text{best}}}$$

- Experimental Data is nice because you can make sure to acquire a large number of observations.
- This paper does not suggest a measure of ”overall completeness” across multiple domains. (no portability consideration)
- Completeness can situationally coincide with the ratio of R-squared of the model class over the non-parametric r-squared.