Comparative Analysis of Machine Learning in Generalized Propensity Score Estimation

Bowen Ma, Min Guo, Yunzhe (Bella) Qian

Department of Biostatistics

Harvard T.H. Chan School of Public Health

Propensity Score

- Definition: The probability of treatment assignment conditional on observed covariates.
- Purpose: approximate the effects of randomization. Balance groups and reduce confoundiness in observational studies.

Data

- Empirical Dataset: Twin
    - US Twin births in 1989-1991. Consists of 52 covariates that are related to infants and their parents, such as parent education and age of mother, and the mortality outcome for both twins after 3 years of birth.
    - Treatment: whether parents smoke. Outcome: whether (≥ 1) child died.
- Monte Carlo Simulated Dataset: Price
    - Consider an airline that wants to use historical data to optimize its ticket prices.
    - Treatment: Price. Outcome: Airline ticket sale performance.
    - Covariate: cost of fuel, time of the year, customer emotion, customer’s demand (high in holiday season)

Binary, Multinomial, Continuous Treatment

- Binary Treatment: Exposed vs. Unexposed
- Multinomial Treatment: Drug A vs. Drug B vs. Drug C
- Continuous Treatment: Dose-response, flight ticket price-decision
- Limitation:
  - Binary comparison not enough;
  - Pseudo-multinomial comparison: step-wise comparison, still binary comparison in each pair, cannot compare simultaneously.
- Contribution: Systematically compare ML propensity score models in binary, multinomial, continuous treatments.

ML Model: Gradient Boosting Machine

Advantage: Non-linear relationship, interaction

Stop method: mean(Kolmogorov-Smirnov statistics) - ks.mean

Optimized tree:

- [Binary/Multi-valued Treatment]

  tree number where minimize ks.mean

- [Continous Treatment]

  tree number where minimize average absolute correlation coefficient

ML Model: Deep Neural Network- Structure

	Binary	Ordinal	Continuous
Input Layer	47	4	4
Hidden Layers	[94, 1175, 293, 94]	[32, 64]	[32, 64, 128, 64]
Output Layer	(1, Sigmoid)	(3, Softmax)	1
Dropout	0.3	0.1	0.3
Loss	Binary CrossEntropy	Sparse CrossEntropy	MSE
Evaluation Metrics	Accuracy+AUC	Accuracy	RMSE+MAE

- Deeper network for more complex dataset (Twin).
- Different set of metrics and loss function for different tasks.

ML Model: Deep Neural Network

Advantage

- Captures complex, non-linear relationships.
- Performs automatic feature representation learning.
- Scalable to data volume and complexity.

Drawbacks

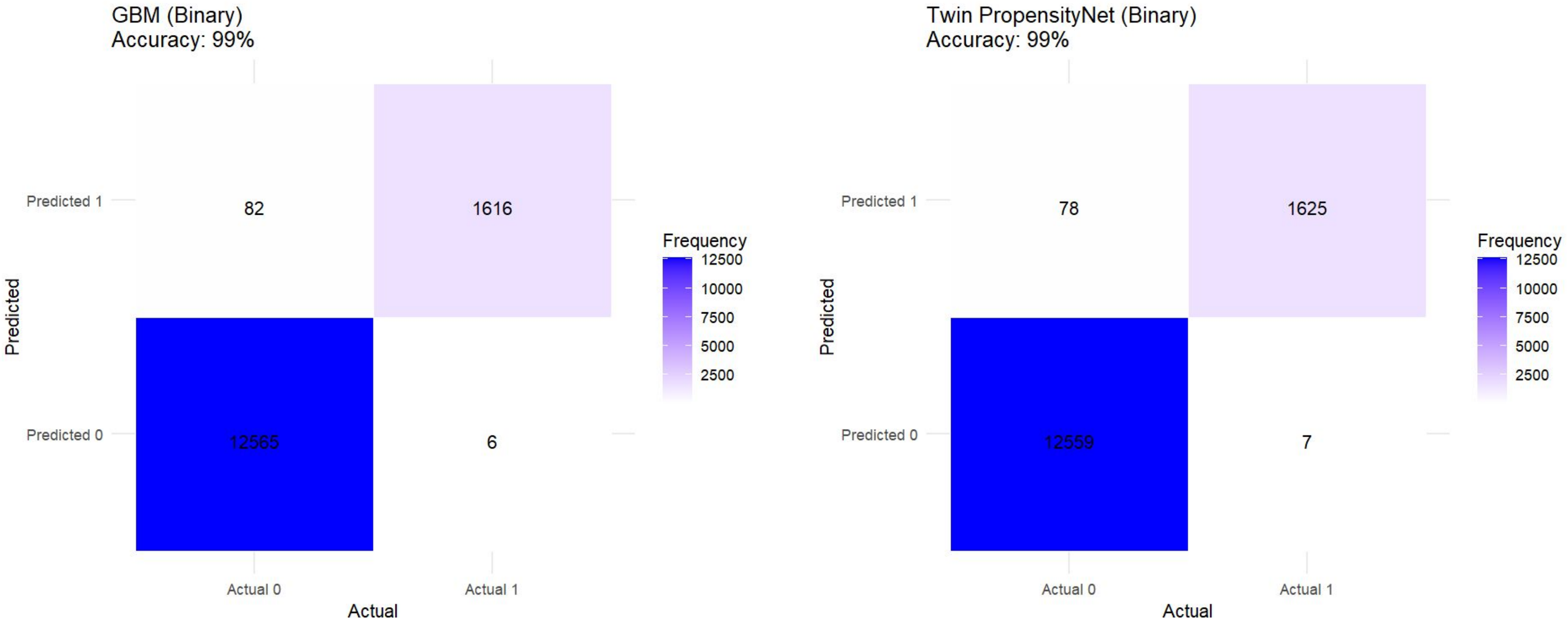
- Prone to overfitting; requires regularization. → L2 Reg + Dropout
- Often a 'black box' with low interpretability. → Main task is not interpret
- Computationally intensive; needs large datasets. → Early stopping criteria

ML Model: Deep Neural Network- Structure

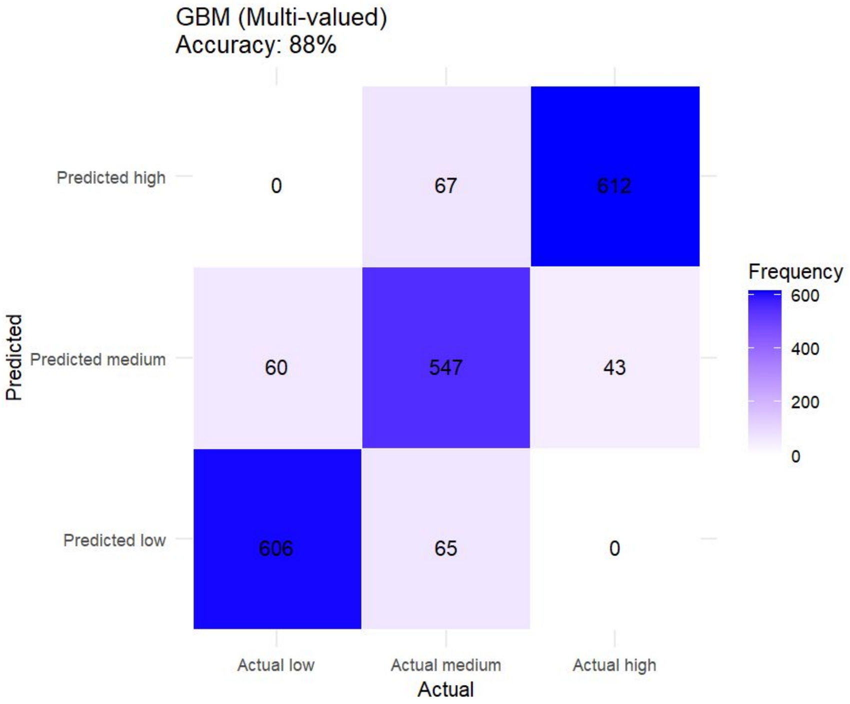
	Binary	Multinomial	Continuous
Input Layer	47	4	4
Hidden Layers	[94, 1175, 293, 94]	[32, 64]	[32, 64, 128, 64]
Output Layer	(1, Sigmoid)	(3, Softmax)	1
Dropout	0.3	0.1	0.3
Loss	Binary CrossEntropy	Sparse CrossEntropy	MSE
Evaluation Metrics	Accuracy+AUC	Accuracy	RMSE+MAE

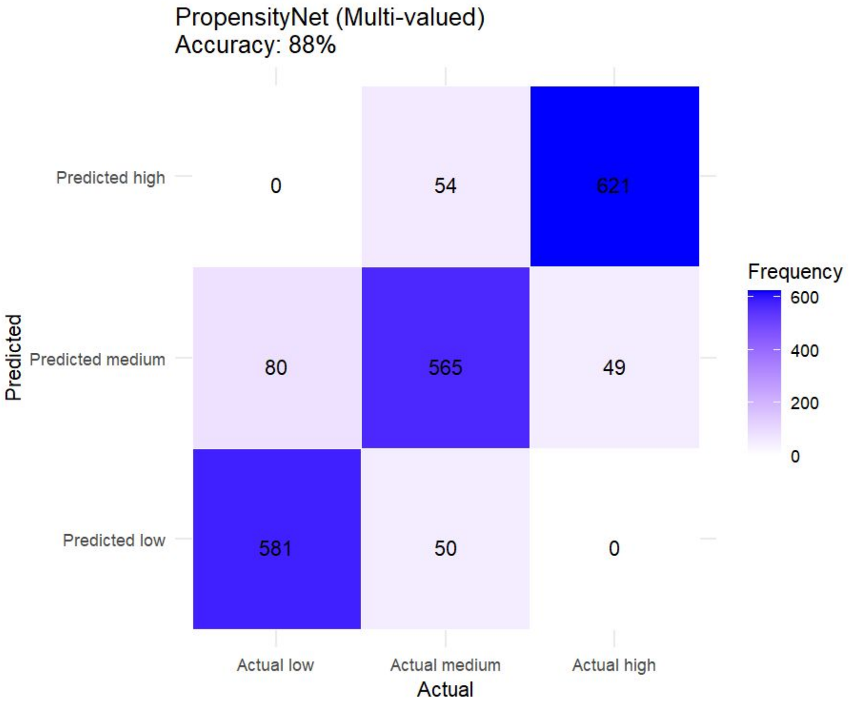
Tasks	Loss	Evaluation Metrics	Results
Binary	0.0364	Accuracy	99.40%
Ordinal	0.2553	Accuracy	88.85%
Continuous	1.9060	RMSE	1.3637

Performance: GBM vs. PropensityNet (Binary)

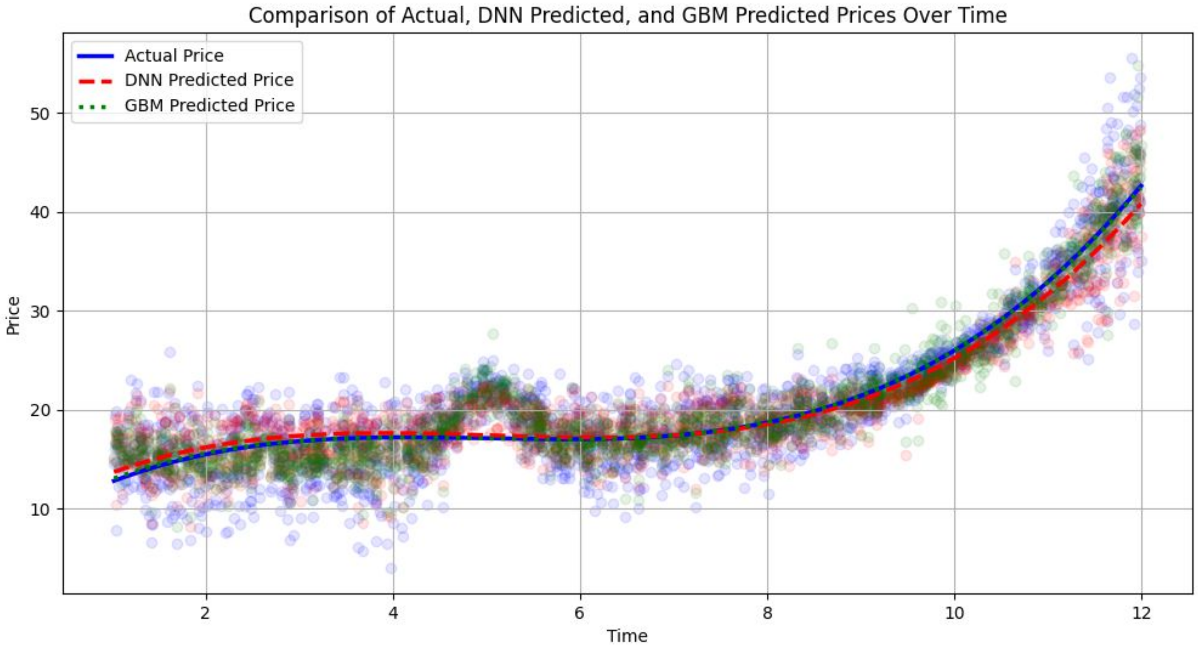


Performance: GBM vs. PropensityNet (Multinomial)





Performance: GBM vs. PropensityNet (Continuous)



	GBM	DNN
RMSE	2.755	1.364
MSE	7.588	1.860