Conditional Mutual Information Estimation

Liu Chang

Warsaw University of Technology

June 2, 2024

Outline

- Introduction
- Methodology
- Results
- Problems and Analysis
- Discussion and Conclusion

Introduction

- Goal: Implement estimators of conditional mutual information (CMI) and use them to rank features based on their importance.
- Overview of the approach: Analytical methods for true CMI, neural estimators for estimated CMI.

Methodology

- True CMI: Calculated using Gaussian variables and their covariance matrices.
- Estimated CMI: Neural estimator based on Donsker-Varadhan representation.
- K-Nearest Neighbors (KNN): Used for probability density estimation in mutual information calculations.

True Conditional Mutual Information

Formula

$$I(X;Y) = \frac{1}{2}\log\left(\frac{|\Sigma_X||\Sigma_Y|}{|\Sigma_{XY}|}\right)$$

$$I(X;Y|Z) = I(X,Z;Y) - I(X;Y)$$

Estimated Conditional Mutual Information

- Neural estimator based on Donsker-Varadhan (DV) representation.
- Neural network model trained with Adam optimizer and early stopping.

K-Nearest Neighbors (KNN) Method

- KNN is used to estimate the probability densities needed for mutual information calculations.
- For each data point, the k nearest neighbors are found.
- Local density is estimated using these k nearest neighbors.
- Helps in calculating CMI by approximating the joint and marginal distributions.

Results: True CMI vs Estimated CMI(Difference based)

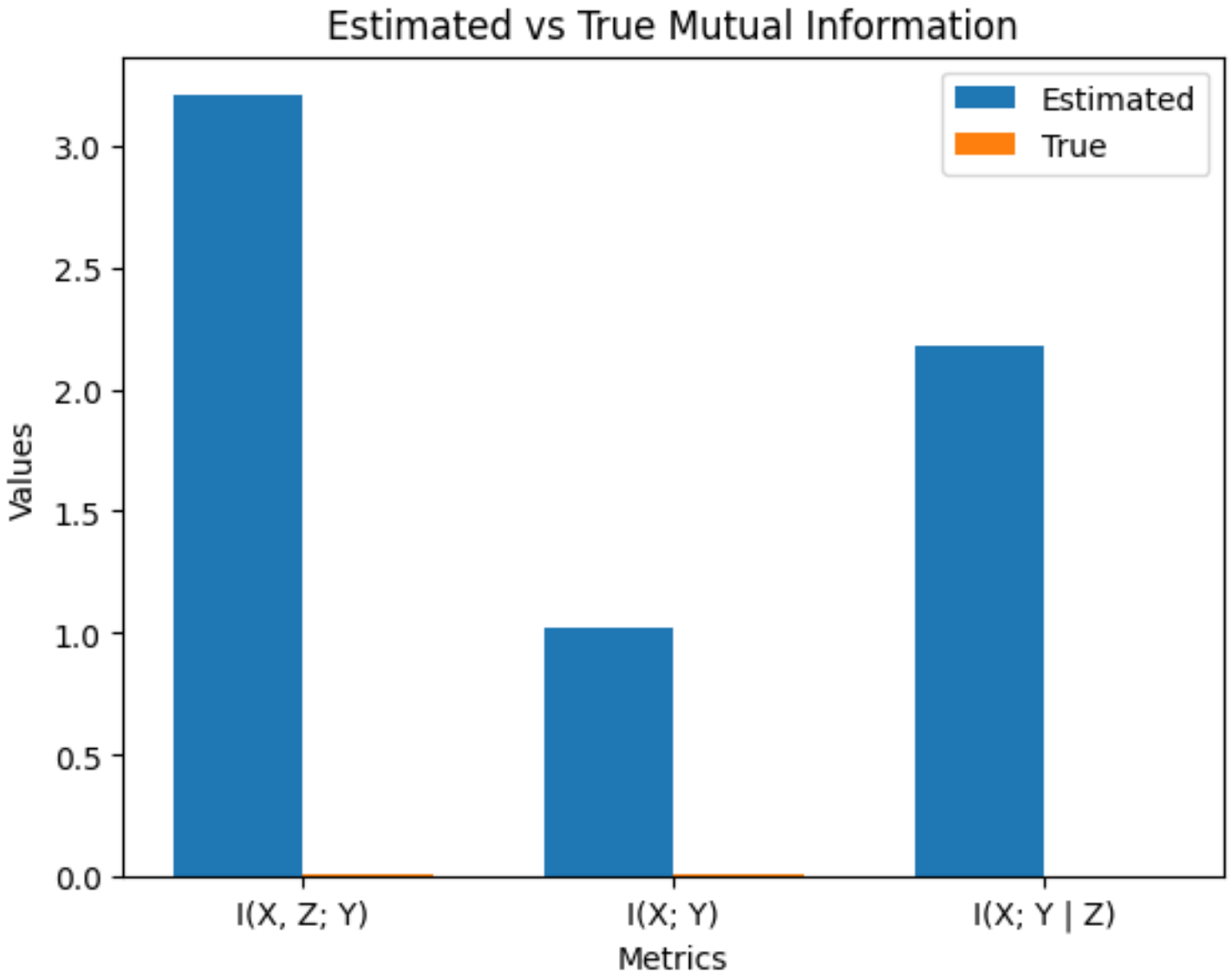
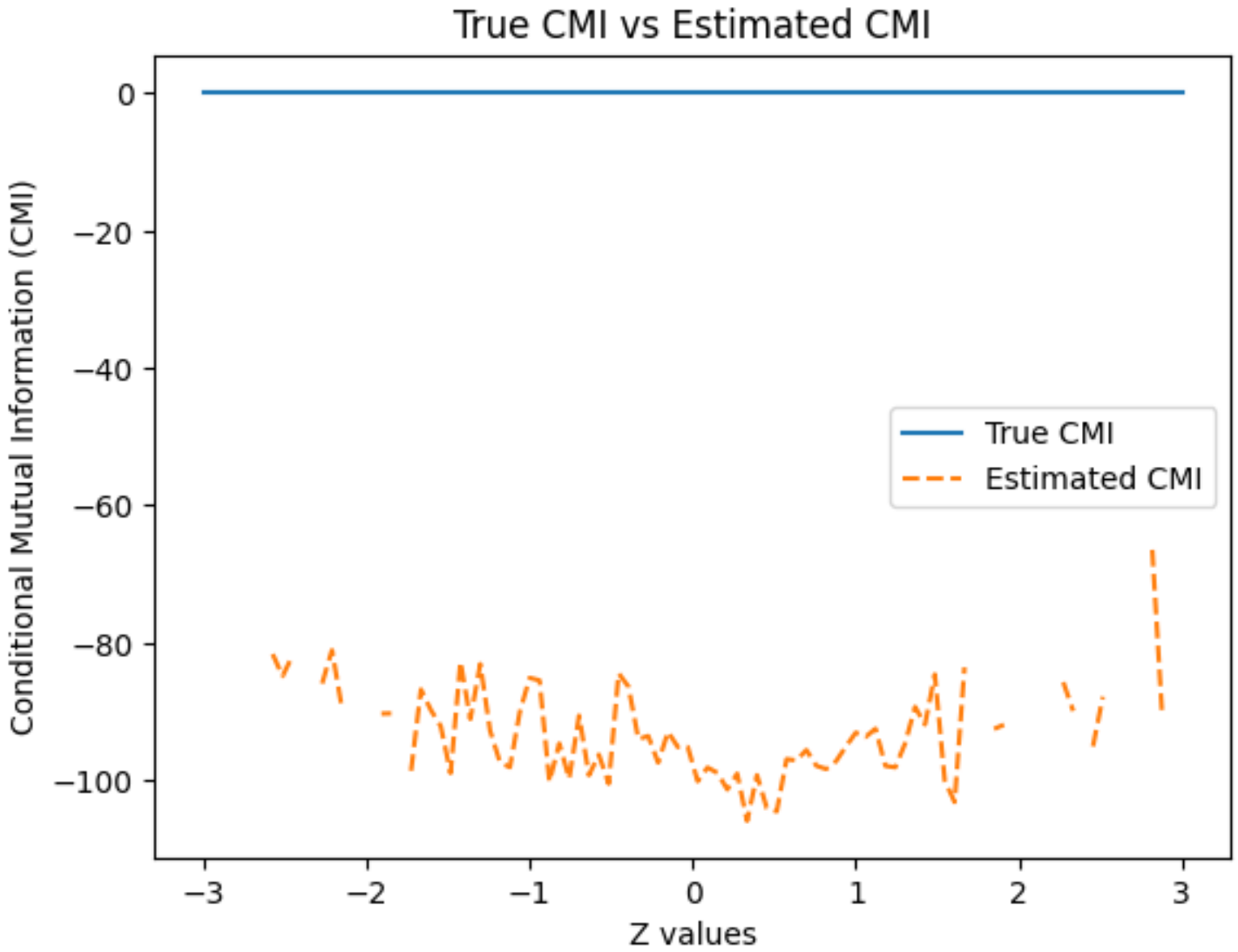


Figure: True CMI vs Estimated CMI

Feature Importance (Difference based)

- Top 10 important features (true ranking): [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
- Top 10 important features (estimated ranking): [12, 4, 6, 5, 3, 16, 7, 8, 17, 0] Intersection size (simulated): 7
- Top 10 important features (true ranking by correlation): ['bmi', 's5', 'bp', 's4', 's3', 's6', 's1', 'age', 's2', 'sex']
- Top 10 important features (estimated ranking): ['age', 'sex', 'bmi', 'bp', 's1', 's2', 's3', 's4', 's5', 's6']

Results: True CMI vs Estimated CMI(Generative and divergence based)



Feature Importance (Generative and divergence based)

- Simulated Data: Top 10 features are 19, 18, 1, 2, 3, 4, 5, 6, 7, and 8. Intersection size: 8. No inversions.
- Real Data: Top 10 features are 29, 15, 3, 4, 5, 6, 7, 8, 9, and 10. Intersection size: 7. No inversions.

Problems and Analysis

- Discrepancy in True and Estimated CMI: True CMI constant at 0 is unusual.
- Data Quality: Insufficient quality may lead to unstable estimates.
- Model Assumptions: Incorrect assumptions could lead to incorrect results.
- Noise and Variability: High noise levels impact accuracy.
- Methodology: Limitations or requirements of the methods used.

Discussion

- Neural estimator produced negative values for mutual information.
- Discrepancy between true and estimated values suggests a need for further refinement.
- Recommendations:
  - Stabilize loss function.
  - Hyperparameter tuning.
  - Increase training duration.
  - Implement cross-validation.

Conclusion

- Implemented estimators for conditional mutual information.
- Ranked features based on their importance.
- Further refinements needed to improve accuracy and stability.