Information-Theoretic Insights for Storage-Limited Learning

IoBT REIGN

Relevant to IOBT CRA Research Area / Task: Research Area 1, Task 1.3

Objectives

- Develop new information-theoretic performance bounds for machine learning algorithms
- Apply bounds to develop improved compression algorithms for storage-limited machine learning
- Support machine learning in battlefield environments with limited storage on mobile terminals

Approach

- Improved bound on generalization error: use mutual information between *individual* training samples and parameters (output hypothesis) of learning algorithm to get a tight bound – Individual Sample Mutual Information (ISMI) bound
- Information-theoretic framework can be used to explain the *population risk* improvement due to model compression

$$\mathrm{E}[L_\mu(\widehat{W})] = \mathrm{E}[L_S(W)] + \mathrm{gen}(\mu, P_{W|S}) + \mathrm{E}[L_S(\widehat{W}) - L_S(W)]$$

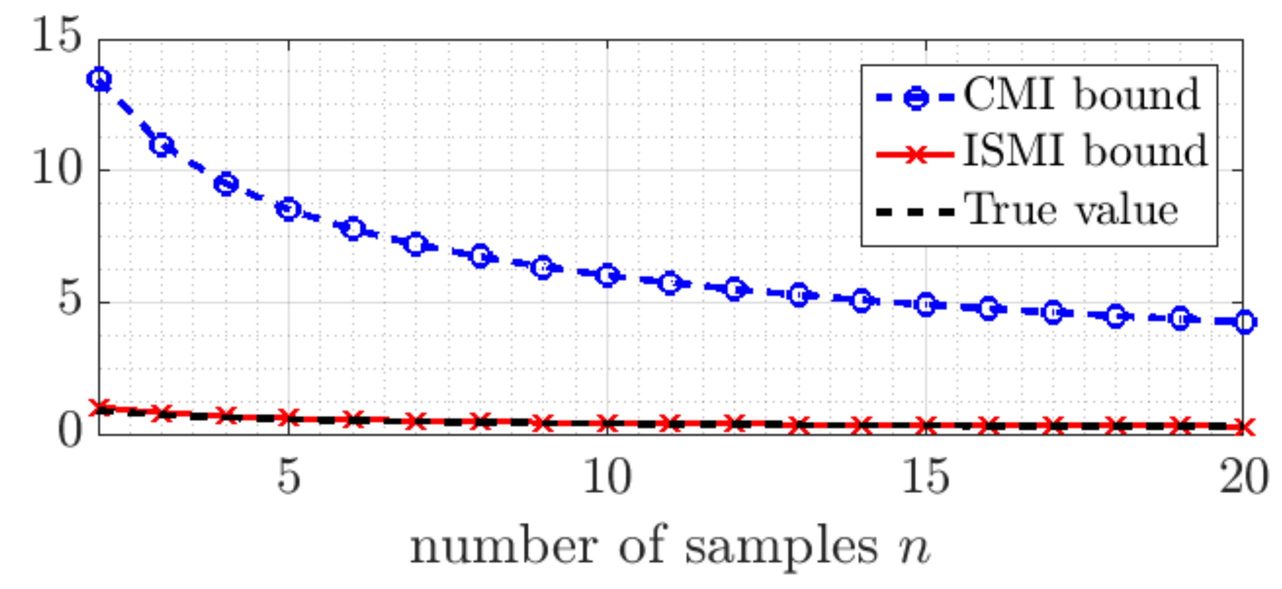
  - Compression increases empirical risk; distortion $\mathrm{E}[L_S(\widehat{W}) - L_S(W)]$ bounded using rate-distortion bound
  - Compression decreases generalization error; $\mathrm{gen}(\mu, P_{W|S})$ bounded using generation error bound
  - Population risk $\mathrm{E}[L_\mu(\widehat{W})]$ can *decrease* with compression

- Information-theoretic framework provides insights for designing improved compression algorithms
  - Algorithms (e.g., Hessian weighted K-Means) can be designed to minimize $\mathrm{E}[L_S(\widehat{W}) - L_S(W)]$
  - Add *regularizer* to optimization to simultaneously control generalization error $\mathrm{gen}(\mu, P_{W|S})$

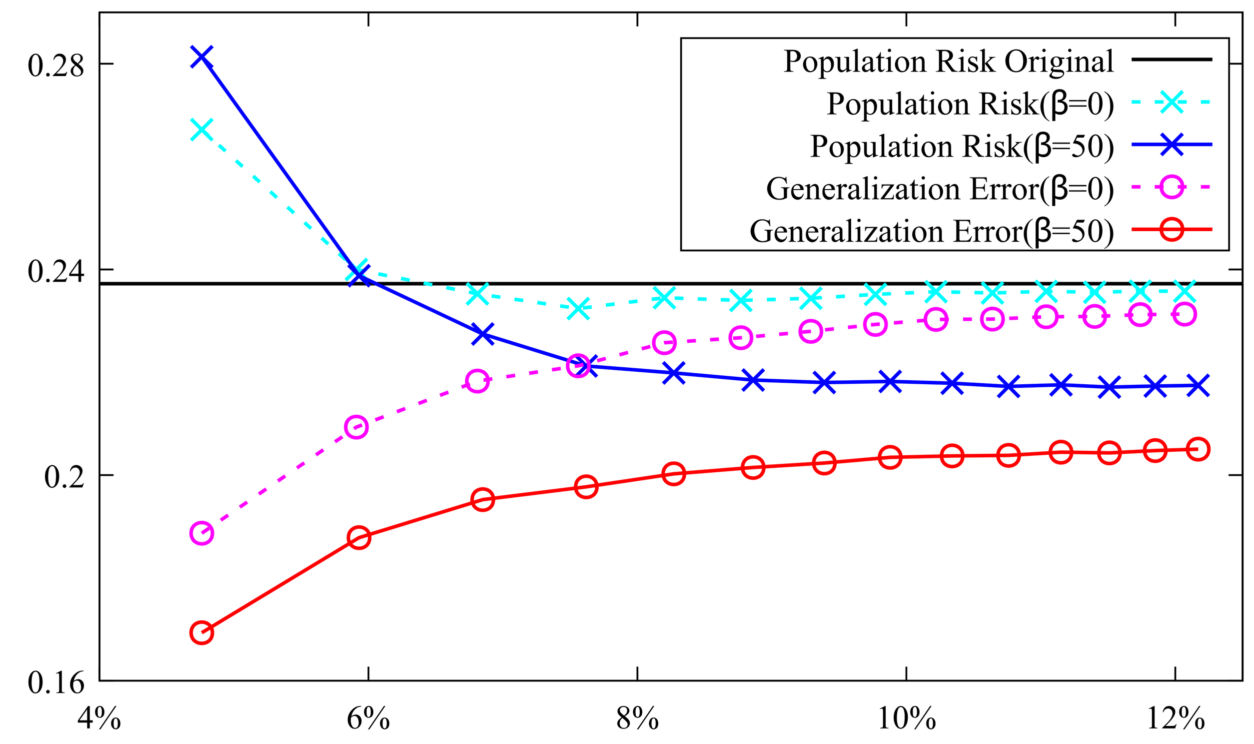
Results

- **ISMI bound on generalization error:**

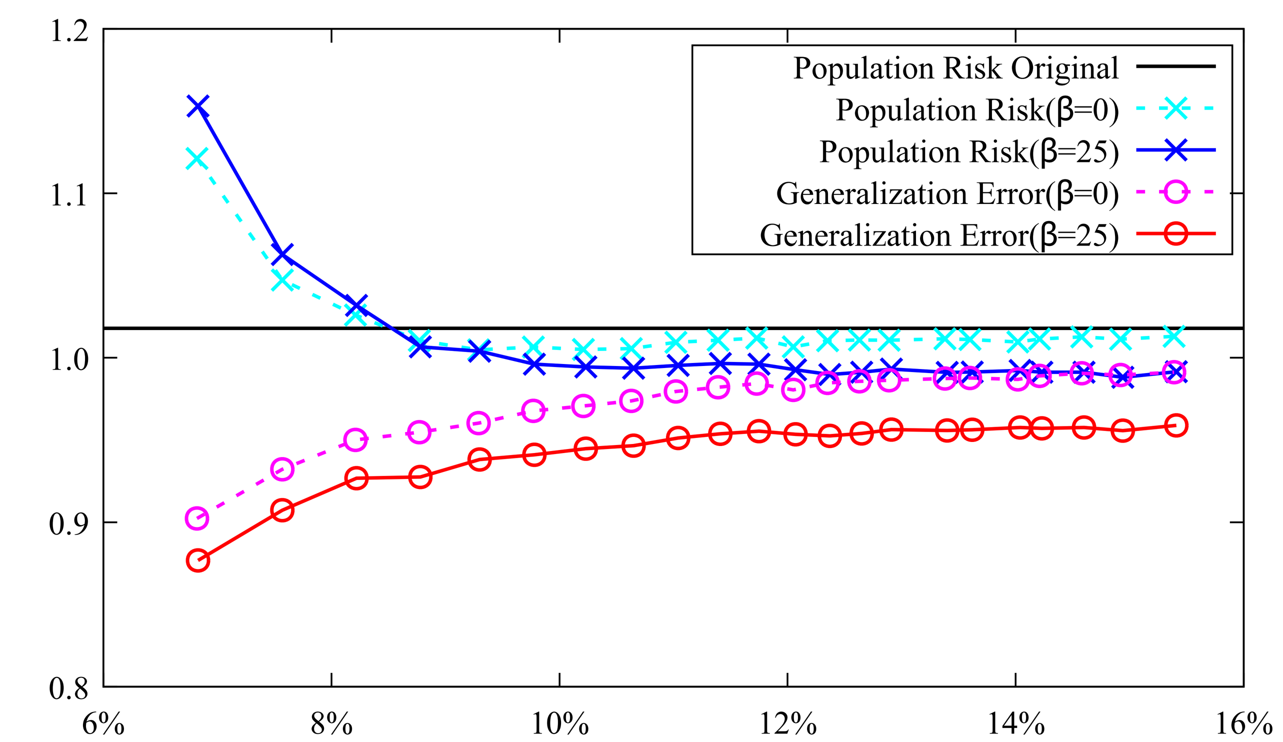
$$|\mathrm{gen}(\mu, P_{W|S})| \le \frac{1}{n}\sum_{i=1}^{n} \psi^{*-1}(I(W; Z_i))$$



- **Compression:**
  - MNIST dataset



  - CIFAR dataset



Conclusions

- Information-theoretic bounds can provide useful insights on performance of machine learning algorithms
- Bounds can be used to improve compression of neural networks – diameter regularized Hessian weighted K-means
- Can compress parameters of neural networks by factor of 15 to 20, without any loss in population risk performance

Path Forward

- Developing information-theoretic generalization bound for case where training data may come from a different distribution than test (population) data to provide robustness
- Developing ways to further improve generalization error through compression, thereby creating more efficient compression algorithms

Publications

- Y. Bu, S. Zou, and V. V. Veeravalli. "Tightening Mutual Information Based Bounds on Generalization Error", in Proc. IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019. Also submitted to Journal on Selected Areas in Information Theory, 2019.
- Y. Bu, W. Gao, S. Zou, and V. V. Veeravalli. "Information-theoretic Understanding of Population Risk Improvement with Model Compression", In Proc. AAAI, New York, NY, 2020.

ILLINOIS USC University of Southern California UMassAmherst UCLA SRI International Georgetown University

POINT OF CONTACT:

Venugopal Veeravalli
vvv@Illinois.edu
217-333-0144