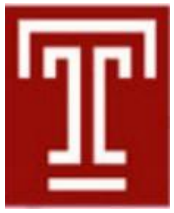


Generalization-Aware Structured Regression towards Balancing Bias and Variance

Martin Pavlovski[1,2], Fang Zhou[1], Nino Arsov[2], Ljupco Kocarev[2], Zoran Obradovic[1]

[1] Center for Data Analytics and Biomedical Informatics
Temple University, Philadelphia, PA, USA

[2] Macedonian Academy of Sciences and Arts
Skopje, Republic of Macedonia

Presented by: Martin Pavlovski (martin.pavlovski@temple.edu)

US-SERBIA & WEST BALKAN DATA SCIENCE WORKSHOP

Belgrade, Serbia | August 26-28, 2018

The Notion of Generalization

- **Intuition:** Striking the proper balance between ***underfitting*** and ***overfitting***

  ⇨ *A fundamental challenge in supervised learning*

Underfitting

- high bias
- Avoided by **reducing** the *empirical risk* R_{emp}

Overfitting

- high variance
- Reduces as the *empirical risk* (training error) becomes a **valid estimate** of the *true unknown risk* (test error):

$$R_{gen} = |R_{emp} - R_{true}|$$

- **Objective:** Minimize R_{emp}, while maintaining low R_{gen}

  ↓ **"measurable"** from the observed data

  ↓ **impossible to determine** since R_{true} is unknown

Main Theoretical Insight

- **Stability-based upper bounds** derived on the *expected true risk* [1, 2]:

$$\underbrace{\hat{R}_{true}(\mathcal{L})}_{\text{Expected true risk}} \leq \underbrace{\mathbb{E}_{\mathcal{D}}[\mathbb{E}_{h|\mathcal{D}}[R_{emp}(h,\mathcal{D})]]}_{\text{Expected empirical risk}} + \underbrace{1 - \mathcal{S}(\ell(\cdot,h), z_{trn})}_{\text{Mutual stability}} \qquad (*)$$

- **A bias-variance balancing objective function**

$$R_{obj}(h,\mathcal{D}) = \sqrt{R_{emp}(h,\mathcal{D})^2 + dCorr(\ell(\cdot,h), z_{trn})^2}$$

- **Aims to tighten the upper bound** $(*)$ by:

  1) **minimizing** the **empirical risk** $R_{emp}(h,\mathcal{D})$

  2) utilizing **distance correlation** [3, 4] to indirectly control the **mutual stability term**

[1] Alabdulmohsin, I. M. "Algorithmic Stability and Uniform Generalization." *NIPS 2015.*
[2] Alabdulmohsin, I. M. "An Information-Theoretic Route from Generalization in Expectation to Generalization in Probability." *AISTATS 2017.*
[3] Székely, et al. "Measuring and testing dependence by correlation of distances." *The annals of statistics 2007.*
[4] Székely, et al. "Brownian distance covariance." *The annals of applied statistics 2009.*

Generalization-Aware Structured Regression

GeneraLization-**A**ware **C**ollaborative **E**nsemble **R**egressor **(GLACER)**

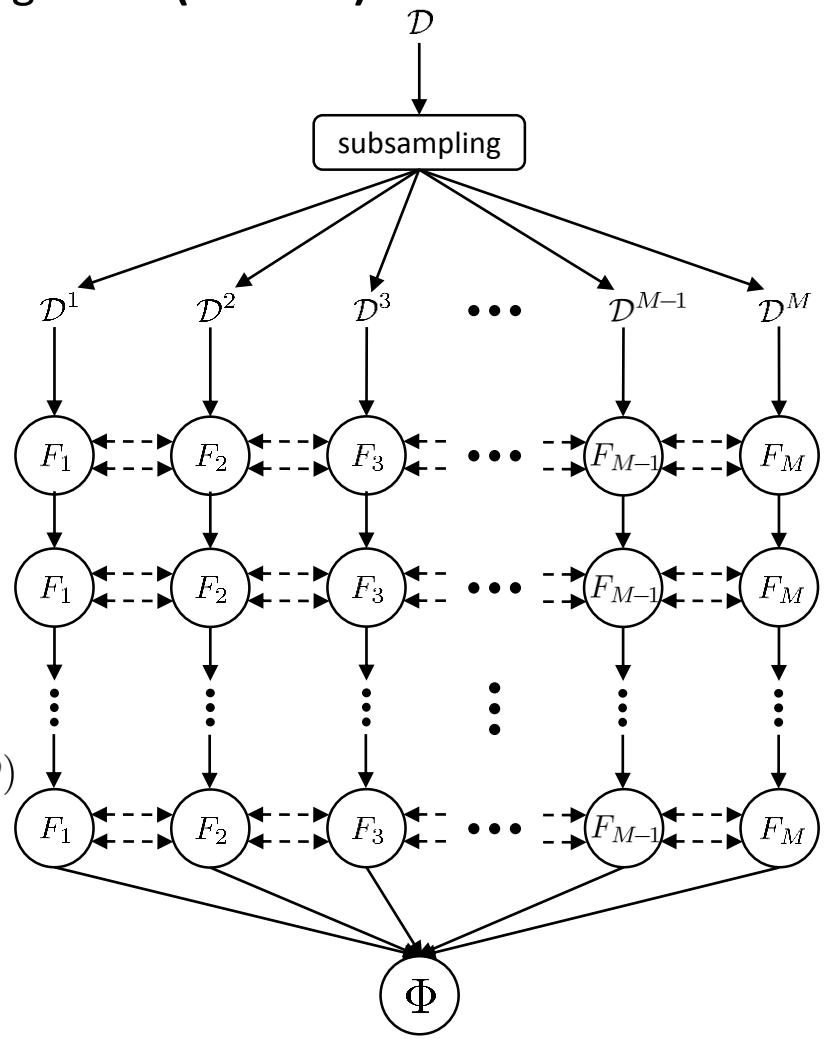
Sample $\mathcal{D}$ and $\mathbf{S}$ (omitted for brevity) M times without replacement using a sub-sampling fraction η

Train a single GCRF component F_m (on top of a least-squares booster) on each $\mathcal{D}^m$ and its corresponding $\mathbf{S}^m$

Loop
- **determine the worst-fit** example for each component
- **exchange** the worst-fit examples between the pair of GCRFs that fosters the highest decrease in $R_{obj}(\Phi, \mathcal{D})$

Repeat until no exchange can further decrease $R_{obj}(\Phi, \mathcal{D})$

Prediction: $\Phi(\mathbf{X}, \mathbf{S}) = \frac{1}{M} \sum_{m=1}^{M} F_m(\mathbf{X}, \mathbf{S})$



Real-World Applications

Sacramento Real-Estate

- Real estate transactions observed in the Greater Sacramento area
- Coupled based on geospatial similarities between houses
- **Task:** predict the housing prices

Medicare Readmissions

- Hospital records from hospitals with more than ~150 readmissions
- Coupled based on similarities between hospital readmissions
- **Task:** predict hospital readmissions

Model	Sacramento	Medicare
Linear Reg.	0.507 ± 0.025	1755.708 ± 616.119
Structured Linear Reg.	0.465 ± 0.024	525.551 ± 196.065
Neural Network	0.516 ± 0.026	2037.421 ± 1199.805
Structured Neural Network	0.463 ± 0.023	1618.547 ± 1192.462
Support Vector Reg.	0.515 ± 0.031	1359.342 ± 697.910
Structured Support Vector Reg.	0.479 ± 0.034	504.076 ± 221.228
Subbagging	0.304 ± 0.017	441.524 ± 101.065
Structured Subbagging	0.262 ± 0.015	234.505 ± 74.378
Random Forest	0.283 ± 0.020	508.294 ± 110.988
Structured Random Forest	0.249 ± 0.015	247.406 ± 35.814
LS Boosting	0.288 ± 0.015	595.289 ± 136.174
Structured LS Boosting	0.250 ± 0.017	182.006 ± 24.919
Convex Network Lasso	0.368 ± 0.013	5012.614 ± 768.945
Non-convex Network Lasso	0.380 ± 0.017	5012.614 ± 768.945
GLACER	$\mathbf{0.225 \pm 0.005}$	$\mathbf{73.183 \pm 9.032}$

Testing MSE, averaged over 10 random splits.

GLACER - Discussion:

- **Outperforms alternatives** by **~10-56% (Sacramento)** and more than **49% (Medicare)**
- Achieves **statistically significant improvements** ⇨ p-values are smaller than 0.01 and 0.021
- Manifests **stable predictions** ⇨ tight confidence interval for its average MSE

Thank you.