Learning Sample-Specific Models with Low-Rank Personalized Regression

Benjamin Lengerich[1], Bryon Aragam[2], Eric P. Xing[1,3]

{blengeri, epxing}@cs.cmu.edu, bryon@chicagobooth.edu

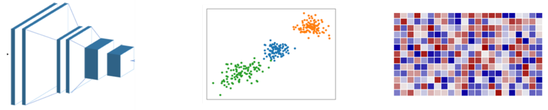
1)Carnegie Mellon University 2) University of Chicago 3) Petuum, Inc.

Guiding Questions

How can we fit machine learning models to help us understand effects which are different for different samples?

Can a **collection** of **simple** models be more accurate than a large model if each simple model is used for only a **single sample**?

Motivation



Deep Learning — Mixture Models — Personalized Regression

Universal, Complicated Effects ⟷ **Personal, Simple Effects**

Can we learn *sample-specific models*? Would give:

- A simple, interpretable model for each sample
- Representational capacity from the entire collection of models.

Let's use a multi-task framework to defining each training sample as a task and learn *sample-specific* models.

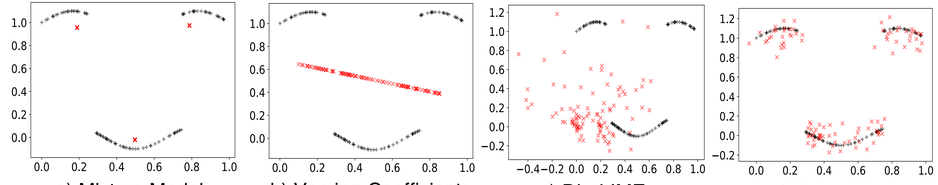
Distance-Matching Regularization

Use **covariates** to induce structure of **parameters**:

$$D_\gamma^i(Z,\phi) = \gamma \sum_{j \in B_r(i)} \Big(\underbrace{\rho_\phi(U^i, U^j)}_{\text{Distance between sample covariates}} - \underbrace{\left\| Z^i - Z^j \right\|^2}_{\text{Distance between sample parameters}} \Big)^2$$

Only uses **distances**, so don't need to prescribe function to transform covariates into regression parameters.

The Benefits of Personalized Models



a) Mixture Model — b) Varying-Coefficients — c) DL+LIME — c) Personalized

Each point represents the regression parameters for a sample: **Black points** are true effect sizes, **red points** are estimates.

- (a) Mixture models estimate a limited number of models.
- (b) The varying-coefficients model estimates sample-specific models but the non-linear structure of the true parameters violates the model assumptions, leading to a poor fit.
- (c) The locally-linear models induced by a deep learning model do not accurately recover the underlying effect sizes.
- (d) Personalized regression accurately recovers effect sizes.

Personalized regression estimates simple models tailored to each sample **without** requiring either: a function to generate regression parameters, or prior knowledge of sample relationships.

Low-Rank Personalized Regression

Each sample becomes a task in a multi-task framework with **soft** parameter sharing:

$$\mathcal{L}(Z,Q,\phi) = \sum_{i=1}^{n} \underbrace{l(X^i, Y^i, Q^T Z^i)}_{\text{Prediction Loss}} + \underbrace{\psi_\lambda(Q^T Z^i)}_{\text{Generic Regularizer}} + \underbrace{D_\gamma^i(Z,\phi)}_{\text{Distance-Matching Regularizer}}$$

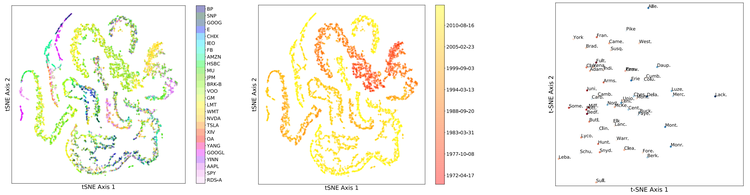
The collection of sample-specific parameters is defined by Z and Q. Dimensionality of Z determines rank. Covariates U determine structure.

Predictive performance

Three datasets in different application areas. Logistic Regressions for classification, Linear Regressions for continuous outputs.

Model	Financial		Cancer		Election	
	R^2	MSE	AUROC	Acc	R^2	MSE
Population	0.01	64144	0.794	0.962	0.00	0.019
Mixture	0.74	16146	0.876	0.939	-0.56	0.031
VC	0.06	60694	0.430	0.863	0.00	0.019
DNN	-0.02	63028	0.901	0.955	0.00	0.019
Personalized	**0.86**	**4822**	**0.923**	**0.975**	**0.45**	**0.011**

Applications: Sample-Specific Models as Embeddings



Predicting Stock Returns. Each point is a t-SNE projection of a regression model for one security at a single date. There is strong clustering in models according to both industry (left) and time (right), but neither covariate would be sufficient to completely characterize each sample.

Predicting election outcomes. Each point is a t-SNE projection of a regression model for one county in Pennsylvania for the 2012 US Presidential Election. Each point is colored according to outcome (blue for Democrat vote proportion, red for Republican vote proportion).

Summary

- We introduce ***Personalized Regression*** to estimate regression models with **sample-specific** parameters.
- Generates a low-rank collection of models with structure that matches the structure of covariate data.
- *Significantly* improves predictive performance on datasets with heterogeneous samples.
- Provides sample-specific interpretability with simple models.