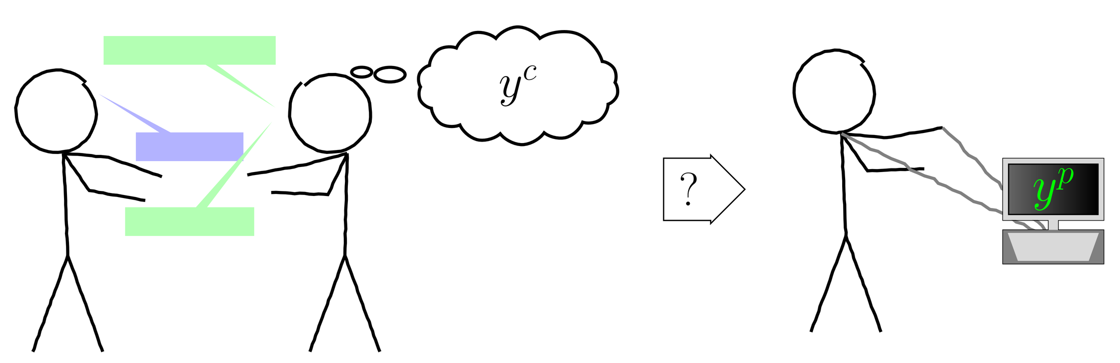
A Sparse Combined Regression-Classification Formulation for Learning a Physiological Alternative to Clinical Post-Traumatic Stress Disorder Scores

Sarah M Brown[1,2], Rami Mangoubi[2], Andrea Webb[2], & Jennifer Dy[1]

[1] Department of Electrical and Computer Engineering, Northeastern University and [2] Software and Algorithms Directorate, Charles Stark Draper Laboratory

Objective

Current diagnosis uses a score derived from clinical interview, the objective is a score computed from physiological measurements



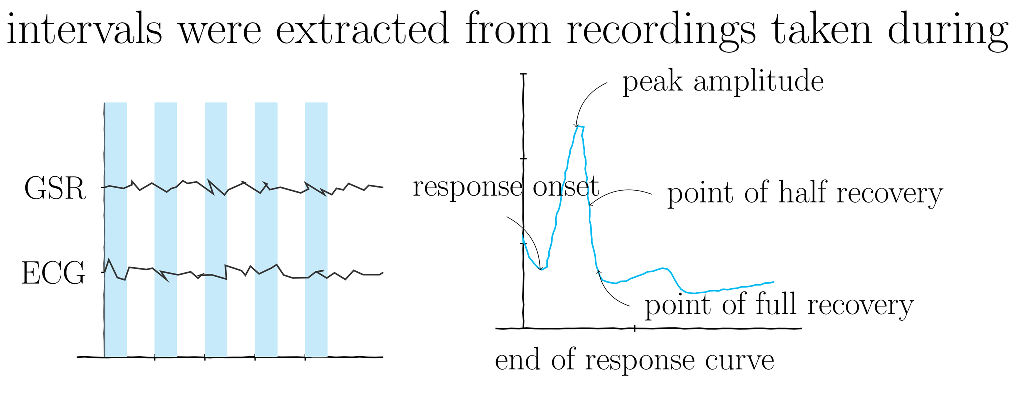
Physiological score must be:

1. Parsimonious- depend on a few features
2. Generalizeable - not overfit to our small dataset
3. Diagnostically Valid- be clinically useful

Data

Experiment: Subjects watch virtual reality videos from Virtual Iraq software, record Electrocardiogram (ECG) and Skin Conductance (GSR) throughout

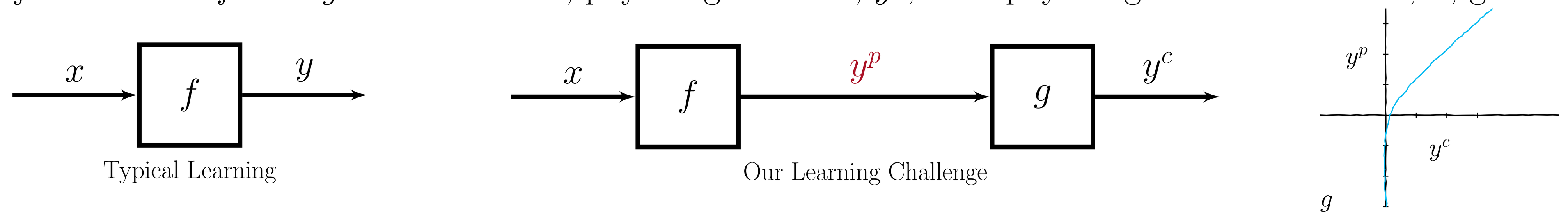
Feature Extraction: For each subject($N = 38$), 5, 20sec response curves at 45 second intervals were extracted from recordings taken during each of the two videos



Abbreviation	Feature Name
AMP	Peak amplitude
LEV	Average level
AFR	Area to full recovery
STD	Standard deviation
AHR	Area to half recovery
FRD	Full recovery duration
FRT	Full recovery time
RR1	Rise Rate from first low point
RR0	Rise Rate from response onset
RT1	Rise time from first low point
RT0	Rise Time from response onset

Application-Tailored Learning

Objective: learn f that *generates* a new, physiological score, y^p, from physiological measurements, x, given measurements and clinical scores, (x, y^c) pairs.



Challenges for developing a learning solution:

- How to formalize expert insight?
- How to learn a scoring function?
- How to measure success?

Learning Desiderata

Expert knowledge about the form of f and g provides a set of desiderata for our learning solution:

1. **Linearity**: Linear with respect to physiological features: $y_i^p = f(\mathbf{x}_i) = \mathbf{x}_i^T\beta$.
2. **Sparsity**: Dependent on only a small subset of the physiological features. Several $\beta_f = 0$.
3. **Severity**: Preserve ranking provided by clinical scores: $y_i^p = g(y_i^c)$, with g nondecreasing $y_i^c > y_j^c \to y_i^p > y_j^p$.
4. **Ambiguity**: Zero scores are non-specific, these subjects present no symptoms, but are not all the same. $y_i^c = 0 \to y_i^p < \epsilon$.

Formulation

From physiological feature data $\mathbf{X}$ and clinical scores $\mathbf{y}^c$:

$$\min_{\beta} \; \gamma_1 \overbrace{\|\mathbf{X}_{\bar{Z}}^T\beta - \mathbf{y}_{\bar{Z}}^c\|_2^2}^{\text{Regression}} + \underbrace{\gamma_2 \sum_{i \in Z \cup \bar{Z}} \mathcal{L}_H(y_i^d, \mathbf{x}_i^T\beta)}_{\text{Classification}} + \lambda \overbrace{\|\beta\|_1}^{\text{Sparsity}}$$

$(i \in Z \cup \bar{Z})$: all subjects

$(\mathbf{X}_{\bar{Z}}, \mathbf{y}_{\bar{Z}}^c)$: Data for subjects with nonzero clinical scores

λ controls sparsity

γ_1, γ_2 control relative weights $(\gamma_1 + \gamma_2 = 1)$ here: $\gamma_1 = \frac{|\bar{Z} \cup Z|}{|\bar{Z} \cup Z| + |\bar{Z}|}$ and $\gamma_2 = \frac{|\bar{Z}|}{|\bar{Z} \cup Z| + |\bar{Z}|}$

$$\mathcal{L}_H(y_i^d, \mathbf{x}_i^T\beta) = \max(0, 1 - y_i^d \mathbf{x}_i^T\beta) \quad y_i^d = \begin{cases} \frac{\max_i y_i^c}{2} & y_i^c > 0 \\ \frac{\max_i y_i^c}{2} & y_i^c = 0 \end{cases}$$

Optimization

We use the alternating direction method of multipliers:

$$\beta_1^{t+1} = (XX^T + \rho I)^{-1}(X^T y + \rho(\beta^t - u_1^t))$$
$$\beta_2^{t+1} = \operatorname*{argmin}_{\beta_2} \sum_i \mathcal{L}_H(y^d, \mathbf{x_i}^T\beta) + \frac{\rho}{2}\|\beta_2 + \beta^t + u_2^t\|_2^2$$
$$\beta^{t+1} = \mathrm{S}_{\frac{\lambda}{\rho}}(.5(\sum_i \beta_i^{t+1} - \sum_i u_i^k))$$
$$u_{1,2}^{t+1} = u_{1,2}^t + \beta_{1,2}^{t+1} + \beta^{t+1}$$

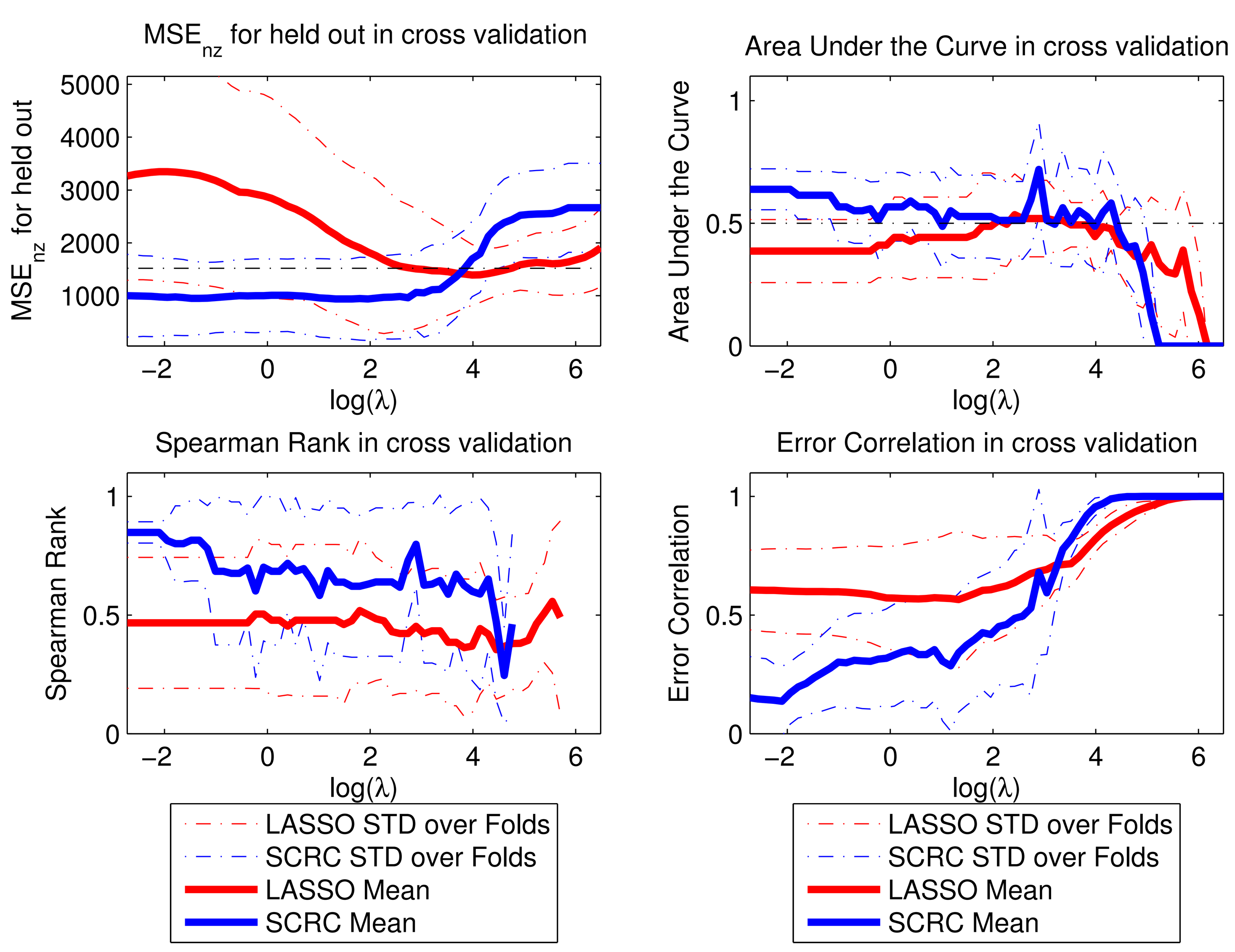
where ρ, the augmented Lagrangian variable, controls how much the difference between solutions regularizes the next iteration and the soft threshold function is defined:

$$\mathrm{S}_\kappa(a) = \max(0, a - \kappa) - \max(0, -x - \kappa)$$

Results: Diagnostic Validity

Diagnostic validity measured through accuracy-type measures in cross validation (7-fold) using the clinical score as ground truth compared to LASSO as a baseline learning method

$$y_i^p = \beta^T \mathbf{x_i}$$
$$A = \bar{Z} \cup \{i; y_i^p > 0\}$$
$$\mathrm{MSE_{NZ}}(\beta) = \frac{1}{N}\sum_{i \in A}(y_i^p - y_i^c)^2$$
$$\rho_S = \mathrm{Lin}(\mathrm{Rank}\,(\mathbf{y}^p), \mathrm{Rank}\,(\mathbf{y}^c))$$
$$\rho_\varepsilon = \mathrm{Lin}(y_i^p - y_i^c, y_i^c)$$



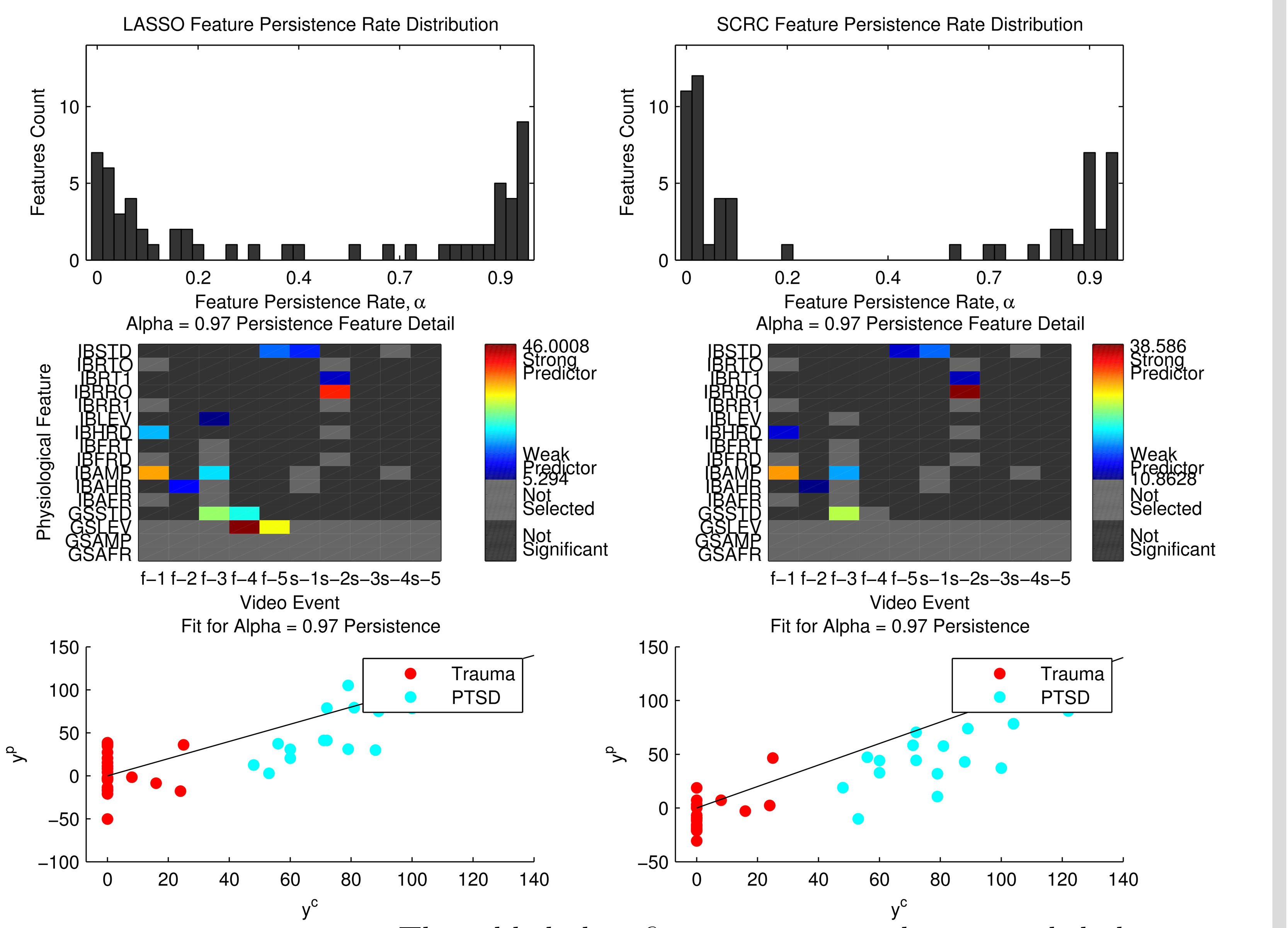
1. SCRC overfits less- it outperforms for small λ (larger number of features)
2. SCRC performance is flat- more robust to overfitting
3. SCRC uses fewer features for a fixed value of λ
4. SCRC error is more uniform across the range of clinical scores

Results: Generalizability & Parsimony

Feature Persistence Rate (FPR) is the percentage of folds a given feature is *active*.

$$\mathrm{FPR}(f, \lambda) = \frac{\mathrm{count}_k(|\beta_{f,k}(\lambda)| > 0)}{N} \qquad \bar{\beta}_f(\lambda, \alpha) = \begin{cases} \frac{1}{K}\sum_k \mathbf{B}_{f,k} & \mathrm{FPR}(f, \lambda) > \alpha \\ 0 & - \end{cases}$$

SCRC provides a more persistent model than LASSO through leave one out cross validation:



Metric	LASSO	SCRC
ρ_S	0.70	0.83
$\mathrm{MSE_{NZ}}$	646.0284	628.0155
AUC	0.85	0.89
Features Count	13	9

The added classification term results in a subtle but important change in features selected and their weights.

Features retained come from both channels and various time points as expected

Northeastern University DRAPER LABORATORY

LaTeX TikZposter