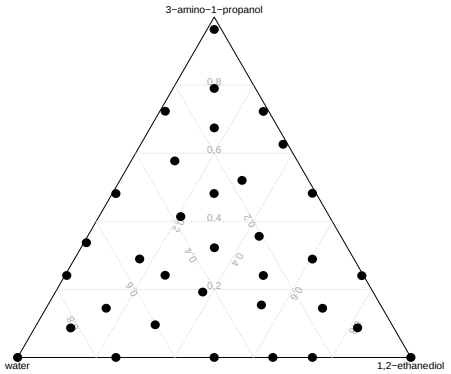
Single-Index Signal Regression

Brian Marx & Bin Li
Department of Experimental Statistics
Louisiana State University

Paul Eilers
Department of Biostatistics
Erasmus Medical Centre
The Netherlands

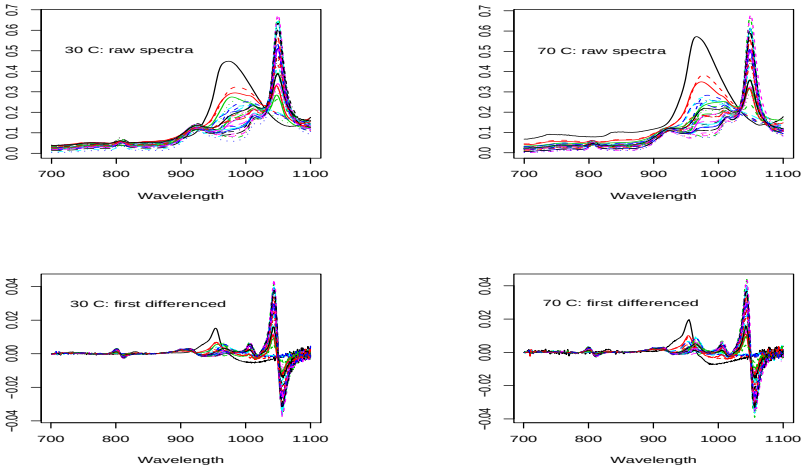
Ternary mixture experiment

- Composition based on mole fraction ($m = 34$)
- Ternary plot: 3 pure $+$ 12 edge $+$ 19 interior



Signal regressors for mixture experiment

- Two (2) temperatures presented (30^o, 70^o)
- Raw and first differenced signals



"Rich" signal regressor information

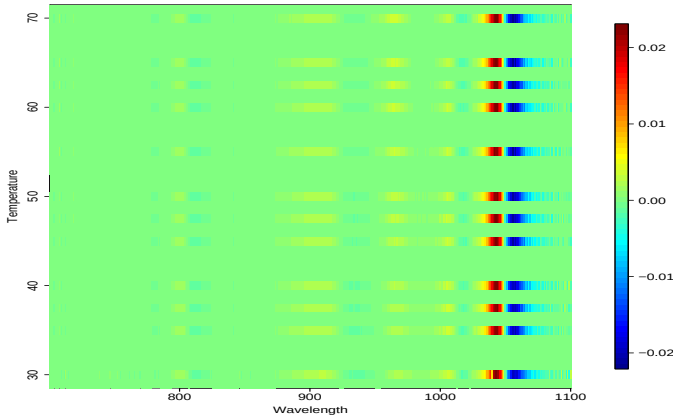
- Digitized optical UV-spectra, $p = 400$
- Ordered structure: 701 to 1100, by 1 nm

- Further: separate spectra at $\breve{p} = 12$ temperatures
- 30, 35, 37.5, 40, 45, 47.5, 50, 55, 60, 62.5, 65, 70^o C

- In total: $p \times \breve{p} = 4800$, for each observation
- Ironic note: we have too much information for OLS
- Many proposals for "regularization"

Exploit "spatial" structure in regressors

- Two-dimensional (first differenced) signal regressor image
- Regressor image for center mixture $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
- Each mixture has unique regressor image



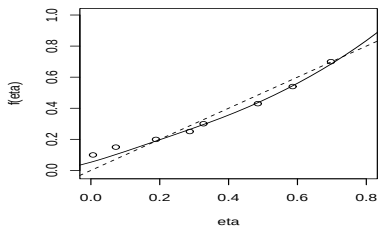
The data structure

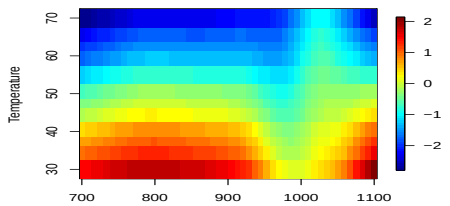
- Each obs, data pair: $(y_i, X_i), \quad i = 1, \ldots, m$
- Responses y: independent, common σ^2
- Regressors X_i (thousands of digitizations)
- (4800) $p \times \breve{p}$ common 2-D array
- Regressor support: $v \times \breve{v}$ (UV-channel, temperature)

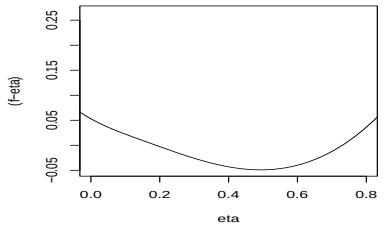
Simultaneous goals

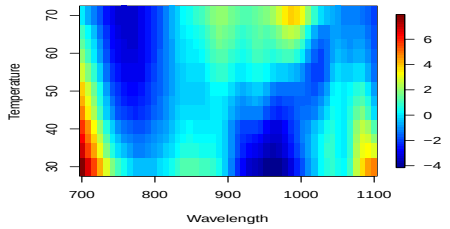
- Parse out two separate modeling components
  1. A single smooth 2D (linear) *coefficient* surface
  2. An unknown *nonlinear* link function
- Explicit modeling of each component
- Nonlinearity could add insight to measurement process
- Combination: systematic, tractable, competitive
- Aim: reliable prediction + interpretability

"Deliverable" for Isopropanol









Model formulation

Standard (penalized) signal regression

$$\mu_i = \sum_{j=1}^{p} \sum_{k=1}^{\breve{p}} x_{ijk}\alpha_{jk}$$

Estimate α as smooth surface

Single-index signal regression

$$\mu_i = f(\sum_{j=1}^{p} \sum_{k=1}^{\breve{p}} x_{ijk}\alpha_{jk})$$

Unknown (univariate) f, with own smoothness penalty

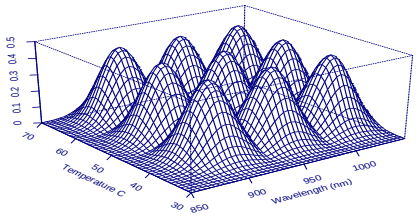
Related to Projection Pursuit: Friedman and Stuetzle (1981)

Signal regression portion

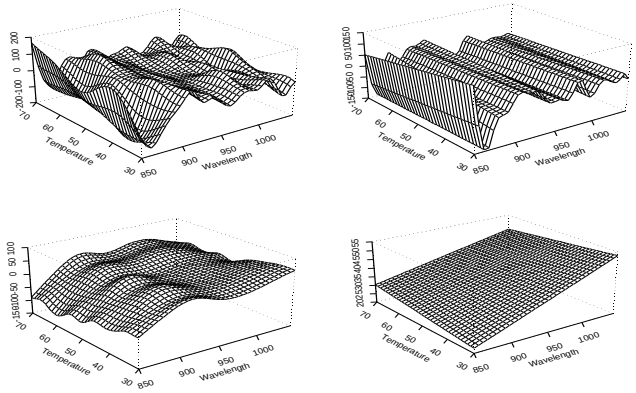
- "Unfold" coefficient surface with tensor product B-splines

$$\begin{aligned} \text{vec}(\alpha) &= \mathbf{T}\gamma \\ \text{vec}(\mu) &= X\alpha = \overbrace{X\,\mathbf{T}}^{\mathbf{M}}\gamma \end{aligned}$$

- (Rich/ Sparse) $\mathbf{T}$: constant γ $\qquad$ $\dim(\mathbf{M}):\ m \times (n\breve{n})$



Examples: (tensor product) *coefficient* surfaces



Penalizing the coefficient surface

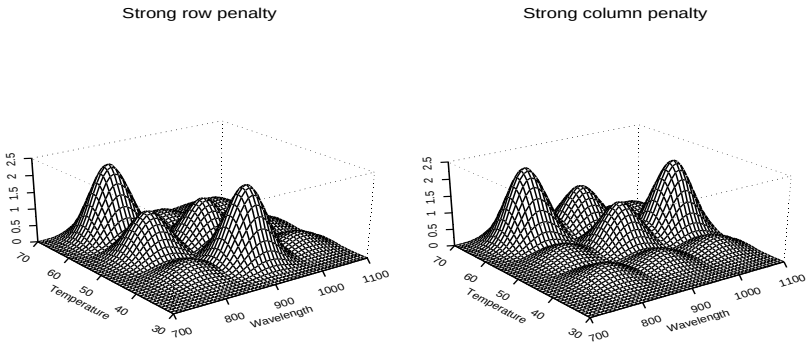
- Minimize:

$$\begin{aligned} Q_P(\gamma) &= \text{Residual SS} + \text{Row Penalty} + \text{Column Penalty} \\ &= ||y - \mathbf{M}\gamma||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2, \end{aligned}$$

- Penalties directly on γ (tensor product coefficients)
- Two positive tuning parameters: $\lambda, \breve{\lambda}$
- Compact representation of penalties:

$$P = (D_d'D_d) \otimes I_{\breve{n}} \quad \text{and} \quad \breve{P} = I_n \otimes (D_{\breve{d}}'D_{\breve{d}})$$

Penalties in "action" (coefficient surface)



Explicit solution:

$$\hat{\gamma} = (\mathbf{M}'\mathbf{M} + \lambda P'P + \breve{\lambda}\breve{P}'\breve{P})^{-1}\mathbf{M}'y$$

Bringing in f

Modified objective:

$$Q_P^\star = ||y - f(\underbrace{\mathbf{M}\gamma}_{\eta})||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 + \lambda_f||D_d\alpha||^2$$

- f one-dimensional (P-spline) smoothing (η, y)
- Rich (univariate) B-spline basis
- f has own λ_f
- Bonus: First derivative $\dot{f}$ easy to compute

Estimation, for fixed f

Approximate:

$$f(\mathbf{M}\gamma) \approx f(\mathbf{M}\gamma_0) + \dot{f}(\mathbf{M}\gamma_0)\mathbf{M}(\gamma - \gamma_0)$$

Objective:

$$\begin{aligned} Q_P^\star &\approx ||y - f(\mathbf{M}\gamma_0) - \dot{f}(\mathbf{M}\gamma_0)\mathbf{M}(\gamma - \gamma_0)||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \\ &= ||y^\star - \mathbf{M}^\star\gamma||^2 + \lambda||P\gamma||^2 + \breve{\lambda}||\breve{P}\gamma||^2 \end{aligned}$$

- Boils down to: PSR$(\mathbf{M}^\star, y^\star, (\lambda, \breve{\lambda}), (D_d, D_{\breve{d}}), (n, \breve{n}))$
- Modified response and regressors:

$$y^\star = y - f(\mathbf{M}\gamma_0) + \dot{f}(\mathbf{M}\gamma_0)\mathbf{M}\gamma_0 \quad \text{and} \quad \mathbf{M}^\star = \text{diag}\{\dot{f}(\mathbf{M}\gamma_0)\}\mathbf{M}$$

Basics of fitting algorithm

- Initialize with $PSR(M, y)$: get $\hat{\gamma}$
- Given $\hat{\gamma}$: get $\hat{f}$ with $S(\hat{\eta}, y)$
- Given $\hat{f}$: get derivative and modified $y^\star$, $\mathbf{M}^\star$
- Update $\hat{\gamma}$: $PSR(\mathbf{M}^\star, y^\star)$
- Regularize norm of $\hat{\gamma}$
- Cycle until convergence

Single-index signal regression algorithm

1. Initializations:
   - Choose the tuning parameter values $(\lambda, \breve{\lambda}, \lambda_f)$ for Steps 1 and 2
   - Choose number of knots $(n, \breve{n}, n_f)$
   - Choose penalty order $(d, \breve{d}, d_f)$
   - Set all tuning parameters to λ_0 for the initial Step 1 (default 10^6)
   - Create $\mathbf{M} = X\mathbf{T}$
   - Calculate $\hat{\gamma} = \text{MPSR}(\mathbf{M}, y, (\lambda_0, \lambda_0), (d, \breve{d}), (n, \breve{d}))$
2. Cycle until convergence of $\hat{\gamma}$'s
   - Estimate $\hat{f}$ and the estimate of the derivative $\dot{f}$ from $S(\mathbf{M}\hat{\gamma}, y, \lambda_f, d_f, n_f)$
   - Obtain $y^\star$ and $\mathbf{M}^\star$
   - Update $\hat{\gamma} = \text{MPSR}(\mathbf{M}^\star, y^\star, (\lambda, \breve{\lambda}), (d, \breve{d}), (n, \breve{n}))$
   - Constrain $\hat{\gamma}/||\hat{\gamma}||$
3. Prediction: $\hat{y}^{new} = \hat{f}(x^{new}\mathbf{T}\hat{\gamma})$

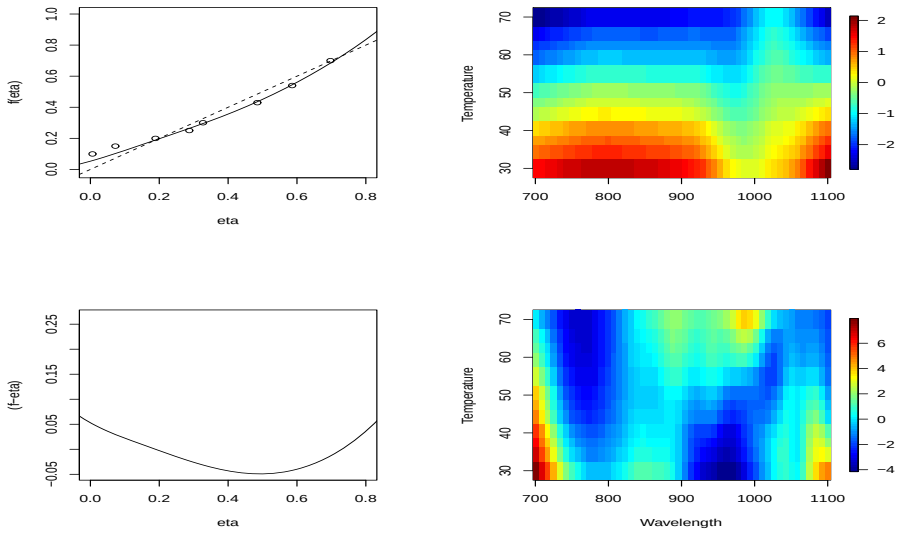
(Optimal) tuning parameters

- Three (3) tuning parameters (λ, $\breve{\lambda}$, λ_f): nm, temp, f
- Data splitting: training, validation, test sets
- Choose λs to minimize:

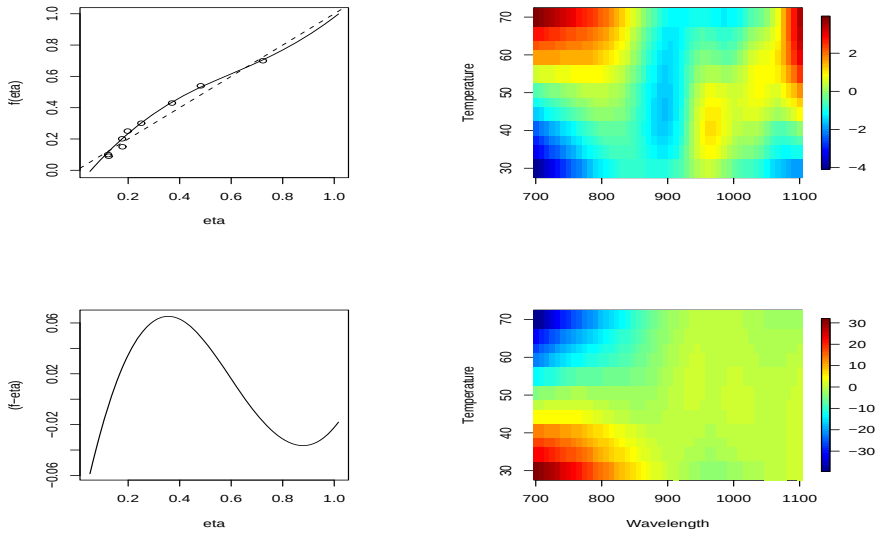
$$\text{RMSEV} = \sqrt{\frac{1}{m^{valid}} \sum_{i=1}^{m^{valid}} (y_i - \hat{y}_{vi})^2}$$

- $\hat{y}_v$: validation prediction, using training parameter estimates

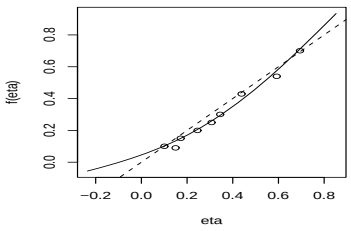
Isopropanol (optimal model)

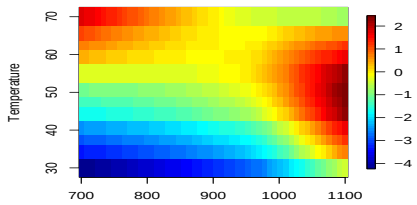


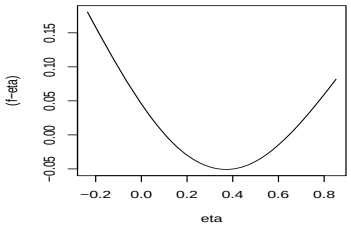
Water (optimal model)

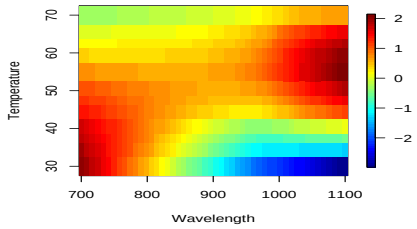


Ethanol (optimal model)









(External) prediction performance

- Given "optimal" model, evaluate

$$\text{RMSEP} = \sqrt{\frac{1}{m^{test}} \sum_{i=1}^{m^{test}} (y_i - \hat{y}_i)^2}$$

- $\hat{y}$: test prediction, using combined (train, valid) estimates
- RMSEP is truly external

Idea of prediction performance: RMSEP

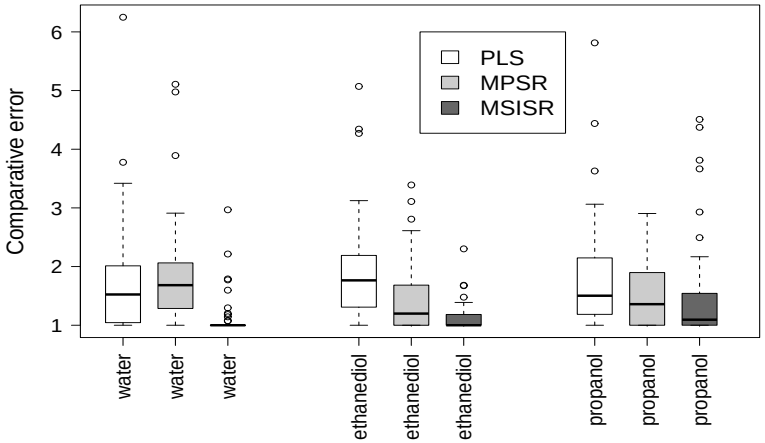
Response	MSISR	MPSR	PLS
Water	0.0214	0.0365	0.0465
1,2-ethanediol	0.0241	0.0338	0.0382
3-amino-1-propanol	0.0306	0.0251	0.0359

- Using optimal models
- RMSEP$\times 100$: units of percent mixture
- MSISR finds reductions of 30% to 55%

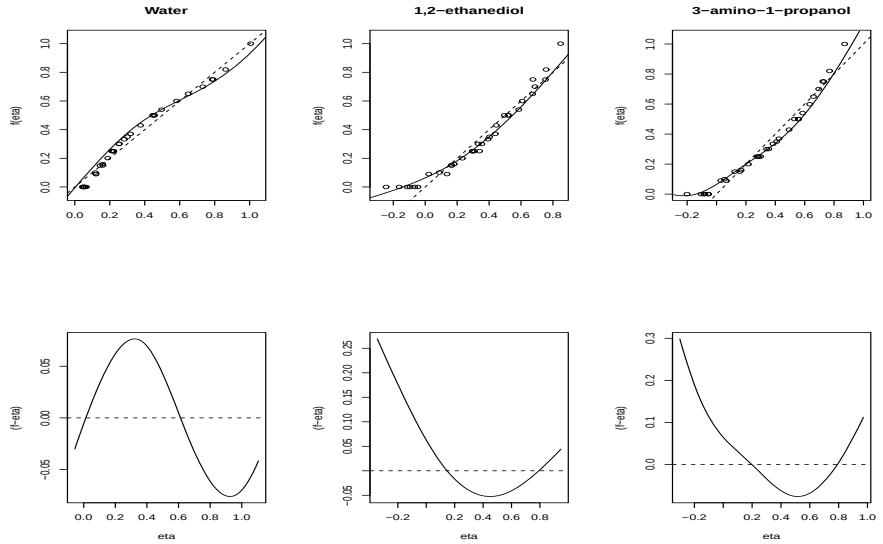
Details: data splitting

- $m = 34$ combinations (16 + 9 +9)
- 16=m^{train} (3 pure + 12 edge + 1 center)
- $9 = m^{valid} = m^{test}$ (each) interior
- Rank 18 interior: use odd (even) for valid (test)
- Fair and reasonable range of mixture levels
- No extrapolation in external prediction

External prediction: 50 random splits

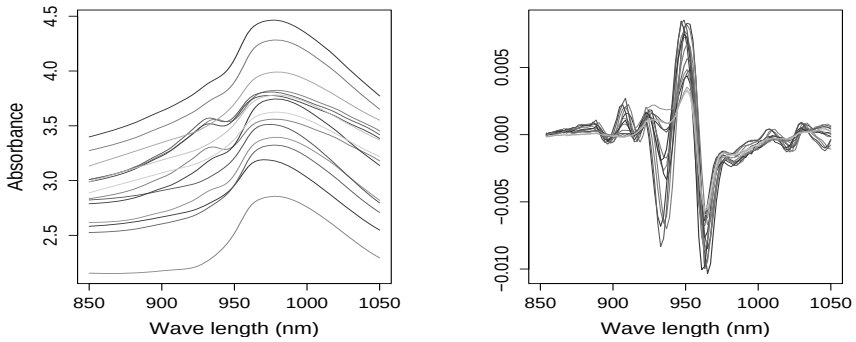


Closer view of f: all mixtures, optimal λs



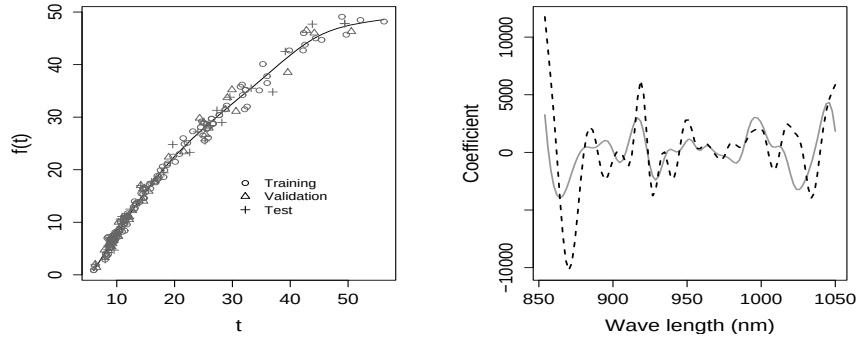
SISR with 1D signals/spectra

- Main differences: 2D **T** is now 1D B; thus no $\breve{\lambda}$
- (Example) Tecator data: $m = 215$: split 129/43/43
- Response: % fat; X contains NIR spectra (100 channel)



Summary results for Tecator experiment

Method	λ	d	n	Valid Error	Test Error
PLS	–	–	–	2.83	2.86
PCR	–	–	–	2.82	2.92
PSR	2.64×10^{-8}	3	100	2.74	1.85
SISR	$(3.36 \times 10^{-6}, 0.616)$	$(3, 2)$	$(100, 20)$	1.73	1.39



Eilers, Li, and Marx (2009), *Chemometrics and Intelligent Lab Sys*

SISR with binary response (GLM)

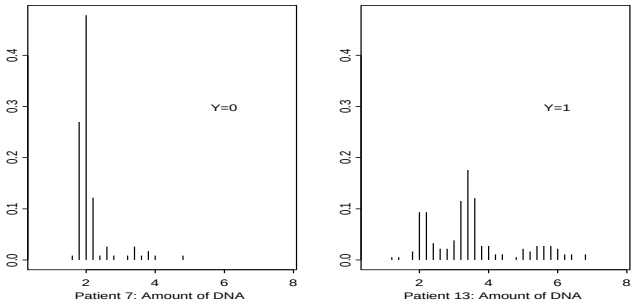
- Now use *iterative* weights: $W = \text{diag}\{\pi(1-\pi)\}$
- Adjusted "dependent variable": $\tilde{y} = W^{-1}(y - \pi) + \eta$
- Cascade functions: $h(f(\eta))$
- h inverse link
- f modifier to h

$$\log\left(\frac{\pi}{1-\pi}\right) = f(\overbrace{\mathbf{M}\alpha}^{\eta})$$

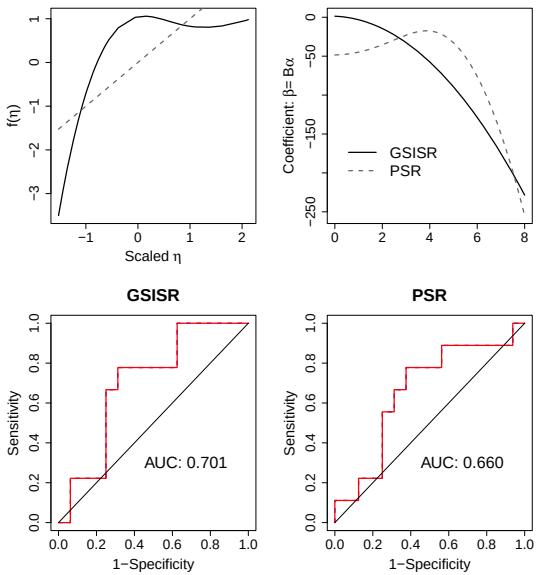
$$0 \equiv \frac{\partial l}{\partial \mu}\frac{\partial \mu}{\partial f}\frac{\partial f}{\partial \eta}\frac{\partial \eta}{\partial \alpha}$$

Ovarian cancer: binary (GLM) example

- $m = 70$ patients (28 "zeros" + 42 "ones")
- Train 65 + Test 25
- Histogram signal (37 DNA bins)



Summary results for DNA experiment



Some benefits of combined approach

1. An explicit (smooth) f can be estimated
2. Heavy penalization: polynomial f
3. Entire signal used as regressors
4. Non-additive coefficient surface allowed
5. Important regions can be identified
6. Allows $p \gg m$
7. Manageable system of equations
8. No "black box"/ No preprocessing
9. Can transplant to generalized linear model