An Autologistic Regression Model for Binary Classification of Hyperspectral Remote Sensing Imagery

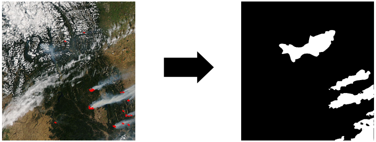
Mark Wolters
Shanghai Center for Mathematical Sciences
Fudan University

Charmaine Dean
Department of Statistical and Actuarial Sciences
Western University

JSM 2016, Chicago
July 31, 2016

What is the talk about?

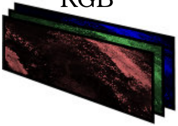
Image segmentation...

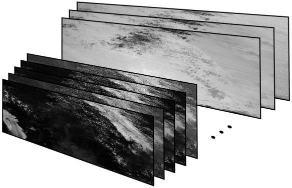


of ***hyperspectral images***...

grayscale

RGB





using ***autologistic regression***.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} + \lambda\,(\text{“autocovariate”})$$

And what did we learn?

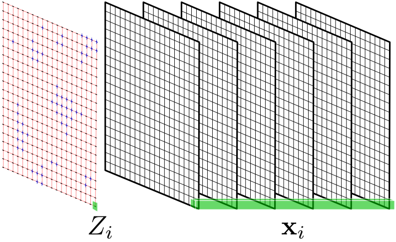
- ***Sequential estimation*** of β and λ is okay.
  - Estimate β first, then find best “plug in” value of λ.

- Autologistic regression may be viewed as ***logistic regression with spatial smoothing***.

- ***–1, 1 coding is essential*** for the procedure to work.

Approaches to image segmentation

- Assume independent pixels.
  - Pixel clustering, many methods
- Spatially-aware methods
  - *ad hoc* rules
  - Model-based: Markov random fields (MRF)
    * Latent random field: $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}) \propto p(\mathbf{X}|\mathbf{Z},\boldsymbol{\theta})p(\mathbf{Z},\boldsymbol{\theta})$
    * Conditional random field: just let $p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})$ be a MRF



Random class label Z_i has covariates $\mathbf{x}_i$

Autologistic regression (1/2)

- Z_i = class label of pixel i.
- Z_i's arranged on a graph.
- $\pi_i = \Pr(Z_i = \text{high} \mid \text{labels of all of } i\text{'s neighbors})$
- ***Standard*** model:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \lambda \sum_{j \sim i} z_j \qquad \text{where} \qquad z \in \{0,1\}$$

- ***Centered*** model (Caragea & Kaiser, 2009):

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \lambda \sum_{j \sim i} (z_j - \mu_j) \qquad \text{where} \qquad \begin{array}{l} z \in \{0,1\} \\ \mu_j = \mathbf{E}[Z_j | \lambda = 0] \end{array}$$

Autologistic regression (2/2)

- ***Proposed*** model:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = 2\left(\mathbf{x}_i^T\boldsymbol{\beta} + \lambda\sum_{j\sim i} z_j\right) \qquad \text{where} \qquad z \in \{-1, 1\}$$

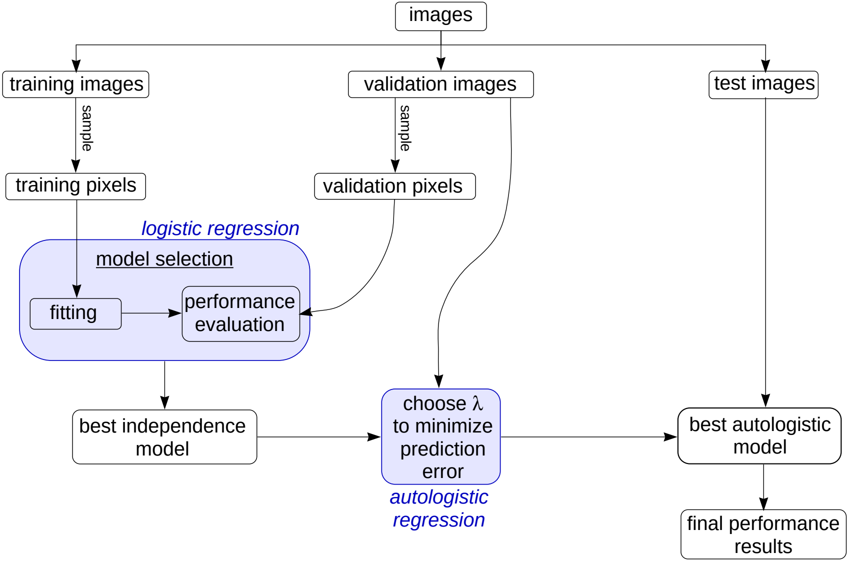
<u>Notes:</u>

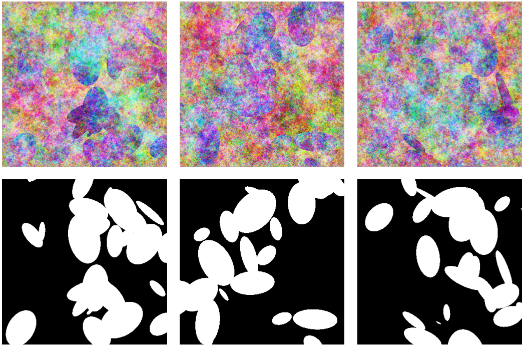
- **Z** is MRF-distributed
  - Conditional, joint PMFs (not shown)
- Normalizing constant intractable
- ***Pseudolikelihood*** for M images, N pixels each:

$$\mathrm{PL}(\boldsymbol{\beta}, \lambda) = \prod_{m=1}^{M}\prod_{i=1}^{N}\pi_i$$

- Our application, $M = 143$, $N > 10^6$, and $\mathbf{x}_i$ is high-dimensional.

Plug-in estimation of λ





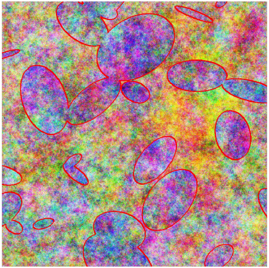
Generating RGB images

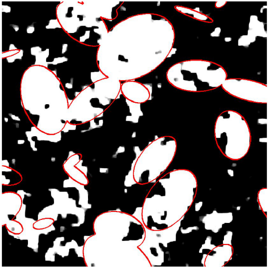
- Random ellipses represent “smoke”
- Color pixels using GMRFs for R, G, B image planes
- Different GMRF parameters for smoke & nonsmoke
- Classes visually overlap
- 90 images each at 100^2, 200^2, 400^2, 600^2, 800^2 pixels

Results on simulated images (2/3)

Parameter estimates and prediction errors:

pixels	method	$\hat{R}$	$\hat{G}$	$\hat{B}$	$\hat{\lambda}$	error rate (%)
100^2	plug-in	-2.21	-2.02	1.91	0.90	20.1
	PL	-2.04	-1.99	2.06	0.99	20.4
200^2	plug-in	-1.64	-1.35	1.71	1.00	17.7
	PL	-1.61	-1.30	1.70	1.19	17.7
400^2	plug-in	-2.05	-1.42	1.63	1.60	20.1
	PL	-2.08	-1.40	1.68	1.36	20.1
600^2	plug-in	-1.91	-1.22	1.76	1.95	20.6
	PL	-1.97	-1.36	1.79	1.51	20.4
800^2	plug-in	-1.55	-1.44	1.58	1.95	18.8
	PL	-1.57	-1.43	1.49	1.59	18.6





Results on simulated images (3/3)

Run time:

pixels	method	time (min)
100^2	plug-in	0.25
	PL	0.49
200^2	plug-in	0.66
	PL	1.5
400^2	plug-in	2.8
	PL	7.5
600^2	plug-in	6.9
	PL	20
800^2	plug-in	12
	PL	35

Computational considerations

- PL bottleneck:
  - Optimization, costly objective function
- Plug-in bottleneck:
  - Sampling to find $\hat{\lambda}$
- If we consider R models?
  - PL: optimize R times
  - Plug-in: $\hat{\beta}$ estimated R times, find $\hat{\lambda}$ once.

Results on the smoke data (1/3)

Logistic model

- Candidate predictors: 35 and 595 interactions
- Logistic GAM approach
  - E.g. for model (2, 3, 4:5),

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + f_2(x_{i2}) + f_3(x_{i3}) + f_{4:5}(x_{i4}x_{i5})$$

  where f's are piecewise linear (5 pieces)

- Model selection
  - Genetic algorithm, model sizes ≤ 18
  - Criterion: validation-set deviance

Autologistic model

- Take best logistic model and plug in λ
- Search for $\hat{\lambda}$ to minimize test-set prediction error

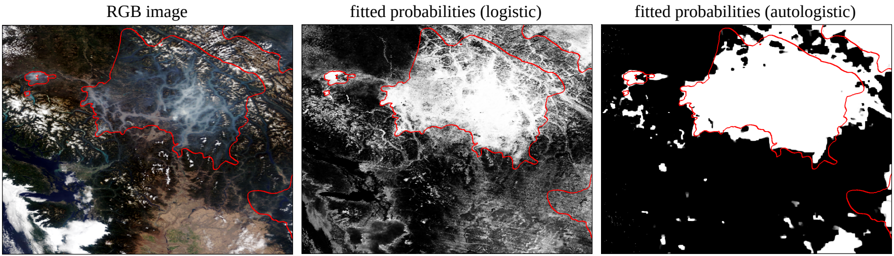
Best models

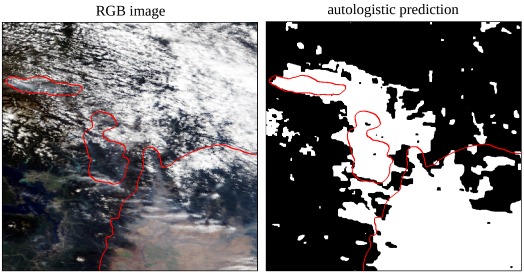
Predictor set	Selected predictors	plug-in $\hat{\lambda}$
main effects	1 6 7 8 14 16 17 18 21 23 25 26 30 31 32 36	1.85
main effects & interactions	7 30 2:3 5:26 6:11 7:36 8:20 8:22 8:25 8:31 13:15 13:23 16:31 18:23 22:36 32:36	1.75

Prediction accuracy

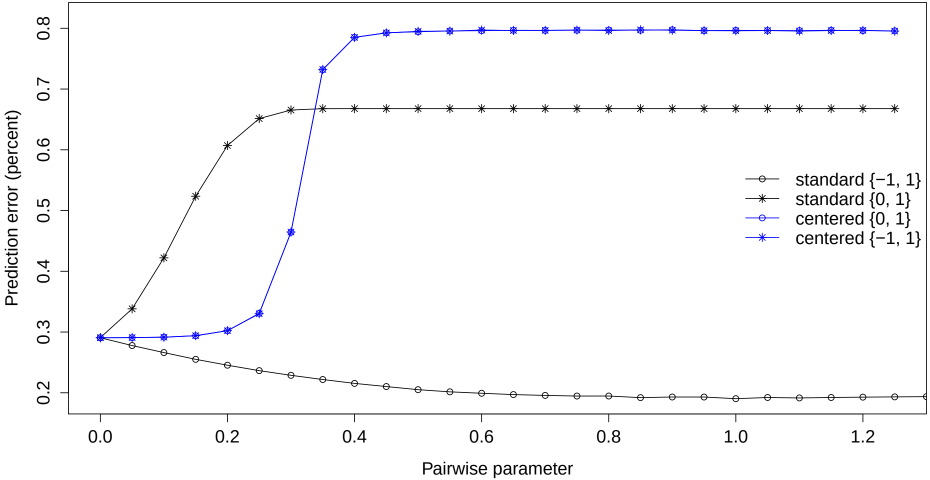
Model	Error rate (%)		
	nonsmoke pixels	smoke pixels	overall
main effects, logistic	21.1	25.9	21.6
interactions, logistic	20.0	23.3	20.3
main effects, autologistic	17.6	23.9	18.2
interactions, autologistic	16.2	21.3	16.7

Results on the smoke data (3/3)





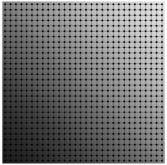
Justifying the plug-in method (1/2)



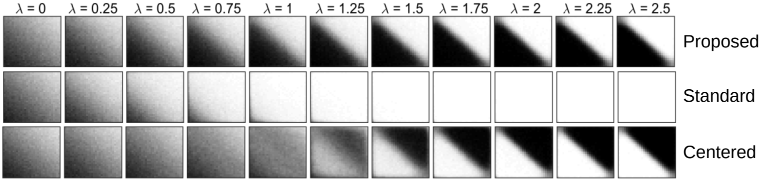
Justifying the plug-in method (2/2)

Facts about autologistic regression models

- Standard, centered, and proposed models are ***not equivalent***.
- Only the proposed model has reasonable behavior as $\lambda \to \infty$.



Independence expectation



Autologistic model's expectation

Conclusions

Summary

- Autologistic regression suitable for binary segmentation
- Model-based segmentation Computationally intensive
- Changing to plus/minus coding enables computational shortcuts
- Using $-1, 1$ coding is a nontrivial change
- Need an adequate independence model for good results.

Acknowledgements

- Data was obtained from the NASA LAADS web portal (`https://ladsweb.nascom.nasa.gov/`)
- The support of NSERC, SharcNet, and Compute Canada is gratefully acknowledged

Related papers

- Wolters & Dean (2015) Issues in the identification of smoke in hyper-. spectral satellite imagery. In book *Current Air Quality Issues*
- Wolters & Dean (2016), Classification of Large-Scale Remote Sensing Images for Automatic Identification of Health Hazards, submitted to *Statistics in Biosciences*
- Wolters (2016), On Coding and Centering in Autologistic Regression, submitted to *JMVA*
- Caragea & Kaiser (2009), Autologistic Models with Interpretable Parameters, *JABES*