Lasso and capture-recapture

Olivier Gimenez and Ian Renner

Introduction

What is this about?

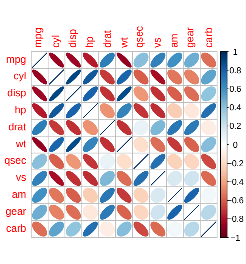
Explain variation in abundance, survival, distribution

With technology, come many variables

Often do not know which ones (not) to include

What are the issues?

Many, possibly correlated, covariates



Correlation $\Longrightarrow$ numerical instability

Many covariates $\Longrightarrow$ $\searrow$ precision and predictability

What we usually do

Think hard about which covariates to consider

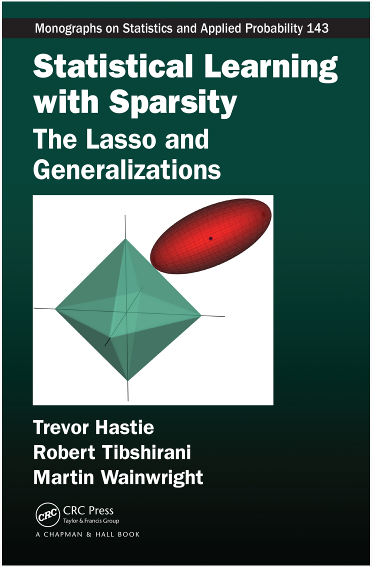
Select covariates using:

- AIC or stepwise procedure
- DIC, SSVS, RJMCMC

This talk: shrink and select model coefficients

Theory

The reference - free book!



It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

$$\widehat{\beta} = \operatorname{argmax} L(\beta) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

It all starts with the ridge regression

Maximize likelihood, penalize magnitude of coeff.

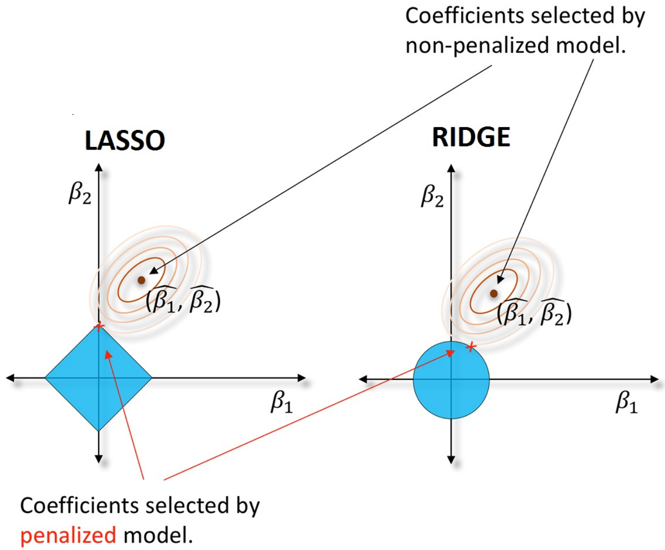
$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} \beta_j^2 < c$$

Lasso = Least Absolute Shrinkage and Selection Operator

Change the constraint: ℓ^2 vs. ℓ^1 norm

$$\widehat{\beta} = \text{argmax}\ L(\beta) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Lasso vs. ridge regression, graphically



Lasso: maximizing penalized likelihood

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Constrained optimization not easy

Rewrite the problem with Lagrange multipliers

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|$$

Lasso: maximizing penalized likelihood

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) \text{ subject to } \sum_{j=1}^{p} |\beta_j| < c$$

Constrained optimization not easy

Rewrite the problem with Lagrange multipliers

$$\widehat{\boldsymbol{\beta}} = \text{argmax } L(\boldsymbol{\beta}) + \lambda \sum_{j=1}^{p} |\beta_j|; \text{ capture-recapt. lik.}$$

How to choose the penalty term λ?

Usually, cross-validation techniques

Build a grid of values for λ

Repeat optimization for each value of the grid

Pick λ corresponding to model with lowest BIC

Application

White storks wintering in Sahel

Capture-recapture data over 16 years

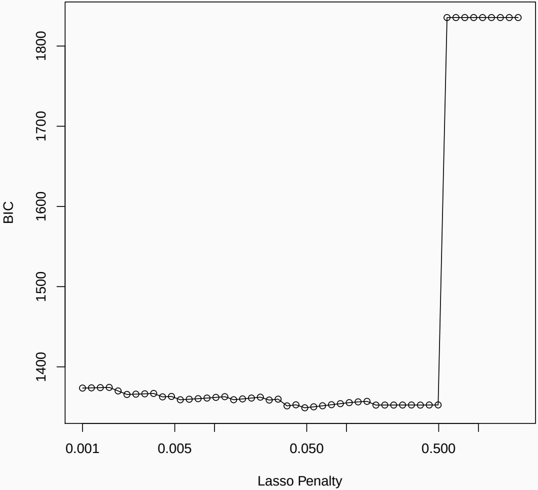
Rainfall was measured at 10 meteo stations in Sahel

Is adult white stork survival affected by rainfall?

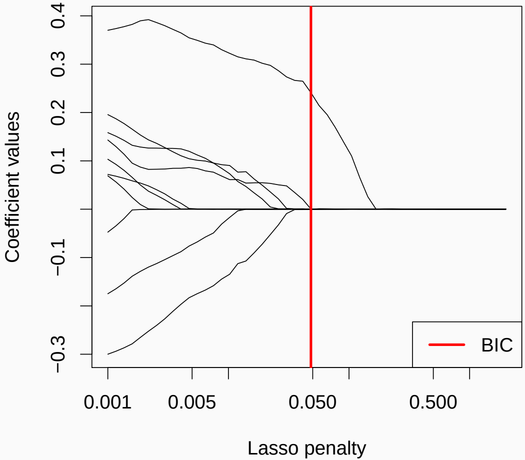
$\text{logit}(\phi_t) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{10} x_{10}$

Do we need to consider 2^{10} candidate models?

Choosing the Lasso penalty using BIC



Exploring regularization path



Rainfall effect at all weather stations

Station	Estimate
Diourbel	7.47×10^{-5}
Gao	-2.99×10^{-5}
Kayes	1.3×10^{-4}
Kita	0.24
Maradi	-1.3×10^{-4}
Mopti	3.5×10^{-4}
Ouahigouya	-5.9×10^{-5}
Segou	1.7×10^{-5}
Tahoua	1.2×10^{-4}
Tombouctou	-2.3×10^{-4}

Almost there

Conclusions and perspectives

From selecting covariates to shrinking estimates

Relatively easy to implement the penalized likelihood

Application to occupancy models

Bayesian flavor, with R. McCrea, E. Matechou and B.J.T. Morgan

Questions

Simulations

Setting: Cormack-Jolly-Seber model

Sample size: 15 occasions, 15 new ind. per occasion

Detection is 0.9, mean survival is 0.8

Covariates: $X_1 \sim N(-0.6, \sigma = 1)$, $X_2 \sim N(0, \sigma = 1)$

Apply Lasso; fit 4 models, compare with AIC

Repeat 100 times

Simulation results

Correct model (X_1 only) is selected 80% with Lasso

Comparable to variable selection using AIC

Further simulations show similar results