STAT 530: More Hierarchical Models

Mar. 31, 2010

Shrinkage revisited

Have seen *unstructured* or exchangeable shrinkage

$\beta_1, \ldots, \beta_m$ conditionally independent *a priori*

So $\hat{\beta}_1, \ldots, \hat{\beta}_m$ shrunk 'toward one another,' compared to 'fitting m separate models.

What about prior judgements whereby 'some β's are more similar than others?

Example

Interested in $E(Y|S)$, with $S \in \{1, \ldots, m\}$

Think of Y_{ij} as response of i-th unit amongst those units with $S = j$.

$$Y_{ij} \sim N(\beta_j, \sigma^2)$$

So $\beta = (\beta_1, \ldots, \beta_m)$ represents $E(Y|S)$.

Smoothness: Want prior with
$\text{Cor}(\beta_j, \beta_{j+1}) > \text{Cor}(\beta_j, \beta_{j+2}) > \text{Cor}(\beta_j, \beta_{j+3})$, etc.

So $\beta_1, \ldots, \beta_m$ conditionally *iid* won't work.

Instead seek hierarchical prior of the form

$$\begin{aligned} \beta|\theta, \lambda^2, \tau^2 &\sim N_m(\mu(\theta), \Sigma(\lambda^2, \tau^2)) \\ p(\theta, \lambda^2, \tau^2) &= p(\theta)p(\lambda^2)p(\tau^2) \end{aligned}$$

One possibility:
$\beta_1 \sim N(\theta, \lambda^2)$
$\beta_2|\beta_1 \sim N(\beta_1, \tau^2)$
...
$\beta_j|\beta_{j-1}, \ldots, \beta_1 \sim N(\beta_{j-1}, \tau^2)$
...

Properties? Role of τ^2

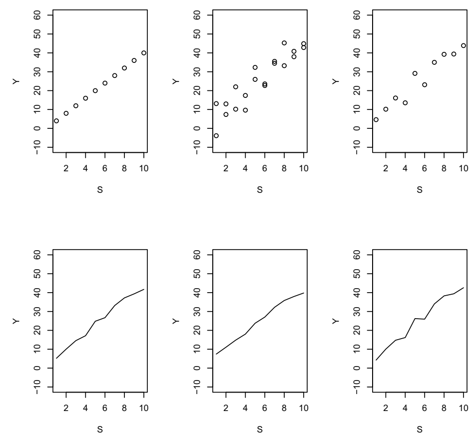
Properties

$\Sigma(\lambda^2, \tau^2) = ?$

$(\beta_j | \text{e.e.}) \sim ?$

Global 'penalization' of 'rougher' functions?

Example (in a further simplified case)



In fact, the model presented is very limited/bad. Can generalize to do much better!

Say S is also continuous, data arise as (S, Y) pairs.

Can model:

$$\begin{aligned} E(Y|S) &\equiv g(S) \\ &= \sum_{j=1}^{m} \beta_j b_j(S) \end{aligned}$$

such that $g(S)$ is a 'cubic spline' on m knots.

And then hierarchical prior

$$\begin{aligned} \beta &\sim N_m(0, \tau^2 V) \\ \tau^2 &\sim p(\tau^2) \end{aligned}$$

with V chosen very specially such that $\int \{g''(s)\}^2 ds = \beta^T V^{-1} \beta$.

Resulting features

Model is very flexible about the form of $E(Y|S)$: very smooth and very wiggly functions of S are both allowed, in principle.

The prior $p(\beta|\tau^2)$ directly encourages/penalizes smooth/rough functions, in an intuitive way.

Via $p(\tau^2)$ and $p(\tau^2|\text{Data})$, the data decide how much smoothing is appropriate.

Important use of Bayes: Not subjective in the sense of prior judgement like $E(Y|S = 7) \approx 3$, etc. Only subjective in the sense that *a priori* I think it more likely that the relationship is smooth, without totally ruling out that it is rough.

Or in a spatial context

If β_j represents effect at j-th spatial location, set up a prior for β reflecting uncertainty about both

- the overall amount of spatial variation
- the smoothness of this variation