Case Study 3: fMRI Prediction

Coping with Large Covariances: Latent Factor Models, Graphical Models, Graphical LASSO

Machine Learning for Big Data
CSE547/STAT548, University of Washington

Emily Fox

February 6th, 2014

Multivariate Normal Models

- So far, we looked at univariate multiple regression

$y^i = \beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i + \epsilon^i \qquad \epsilon^i \sim N(0,\sigma^2) \qquad y^i \in \mathbb{R}$

$= \beta^T x^i + \epsilon^i$

$\Rightarrow y^i \sim N(\beta^T x^i, \sigma^2)$

- If one has a multivariate response $y^i \in \mathbb{R}^d$ ← # of semantic feature
  - Assuming independence between dimensions

$$y^i \sim N\left(\underbrace{\begin{bmatrix} - \beta^{(1)T} - \\ - \beta^{(2)T} - \\ \vdots \\ - \beta^{(d)T} - \end{bmatrix}}_{B^T} x^i, \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix} \right)$$

$\beta^{(l)}$ are reg coeff. for the l^{th} dim

Multivariate Normal Models

- If one has a multivariate response $y^i \in \mathbb{R}^d$ ← "dog" and "furry"
  - Assuming correlation between the output dimensions

  $y^i \sim N(B^T x^i, \Sigma)$

  recall: $\mathrm{cov}(y_s, y_t) = \Sigma_{st}$

- Assume linear (or other mean regression) is removed and focus on the correlation structure

  $y^i \sim N(0, \Sigma)$ ← focus Σ sym., pos. def.

- Matrix valued parameter!

  See more of this in Case Study 4

High-Dimensional Covariance

- What if *d* is large? $\Sigma \in \mathbb{R}^{d \times d}$ s.t. sym., pos. def.

  \# params $(\Sigma) = \frac{d(d+1)}{2}$

  Again, consider $d >> N$
  but $O(d^2)$ params. to est.

- A few common approaches:
  - Low-rank approximations
  - Sparsity assumptions

Low-Rank Approximations

- In general, assume some matrix parameter

$$\underset{d\times m}{\Theta} = \underset{d\times k}{A}\,\underset{m\times k}{B^T} \qquad k \ll d, m$$

will see this in case study 4

- Here, Σ must be a symmetric, positive definite matrix

square

$$\underset{d\times d}{\Sigma} = \underset{d\times k}{\Lambda}\Lambda^T + \Sigma_0 \qquad \Sigma_0 = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}$$

sym. + square

pos. def.

©Emily Fox 2014 5

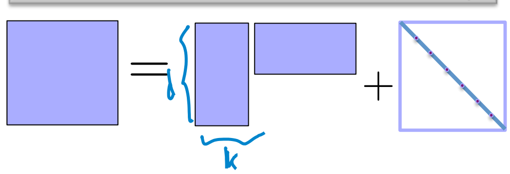
Low-Rank Approximations

- In pictures…

$\Sigma_0 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$

$$\Sigma = \Lambda\Lambda' + \Sigma_0$$

k << d



- Number of parameters:

$$d \cdot k + d = d(k+1) \ll \frac{d(d+1)}{2}$$

sig. reduction in param. for k << d

©Emily Fox 2014 6

Latent Factor Models

- Original multivariate regression

here: assume linear term is removed

$$\mathbf{y}^i = \cancel{B^T x^i} + \epsilon^i, \qquad \epsilon^i \sim N(0, \Sigma)$$

- Latent factor model assumption: $\Sigma = \Lambda\Lambda' + \Sigma_0$
- Low-rank approximation arises from a latent factor model

$$\underbrace{y^i}_{\text{obs}} = \underbrace{\overset{d \times k}{\Lambda}}_{\text{"factor loadings"}} \underbrace{\eta^i}_{\text{latent factors}} + \tilde{\epsilon}^i \qquad \eta^i \overset{iid}{\sim} N_k(0, I) \qquad \tilde{\epsilon}^i \overset{iid}{\sim} N_d(0, \Sigma_0)$$

- Proof:

$$\text{cov}(y; \Lambda, \Sigma_0) = E[(y - \underbrace{E[y]}_{0})(y - \underbrace{E[y]}_{0})^T] = E[yy^T]$$

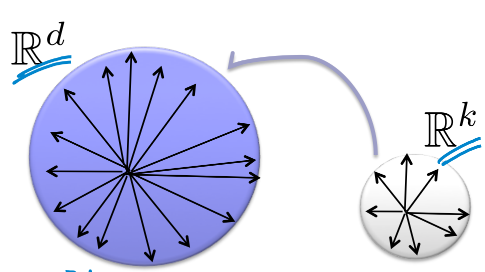
$$= E[(\Lambda\eta + \tilde{\epsilon})(\Lambda\eta + \tilde{\epsilon})^T] = \Lambda \underbrace{E[\eta\eta^T]}_{I}\Lambda^T + 2\underbrace{E[\eta^T]}_{0}\Lambda E[\tilde{\epsilon}] + \underbrace{E[\tilde{\epsilon}\tilde{\epsilon}^T]}_{\Sigma_0}$$

$$= \Lambda\Lambda^T + \Sigma_0$$

Lower-dim Embeddings

Sharing information in
low-dim subspace

obs. y^i



η^i latent factors

"can you hold it?" and "is it bigger than a bread box" } redundant ← latent factor "is it big?"

Sparsity Assumptions

- What if we assume Σ is sparse?

$(i \neq j)$ $\Sigma_{ij} = 0 \Rightarrow y_i \perp\!\!\!\perp y_j$ ← assumin $y \sim N$

$cov(y_i, y_j) = 0$

Could assume Σ sparse to reduce # params,
but each 0 encodes an indep. assumpt...
often too strong

- More often, we can reasonably make statements about *conditional independence*

"cat" $\perp\!\!\!\perp$ "dog" | "animal", "furry", "pet", ...