Case Study 3: fMRI Prediction

Coping with Large Covariances: Latent Factor Models

Machine Learning for Big Data
CSE547/STAT548, University of Washington
Emily Fox
April 30th, 2015

Multivariate Normal Models

- So far, we looked at univariate multiple regression

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

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

so far

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

$\beta^{(l)}$ are coeff. for the l^{th} semantic feature

Multivariate Normal Models

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

"dog" and "furry"

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

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

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

$$y^i \sim N(0, \Sigma) \qquad \Sigma \text{ sym. pos. def.}$$

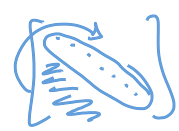
- Matrix valued parameter!

see more on matrix valued params in Case Study 4

High-Dimensional Covariance

- What if d is large?

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



sym

$\Sigma \in \mathbb{R}^{d \times d}$

Again, consider $d >> N$

but $O(d^2)$ params to est.

- A few common approaches:
  - Low-rank approximations ← this lecture
  - Sparsity assumptions

Low-Rank Approximations

- In general, assume some matrix parameter

$$\underset{d\times m}{\Theta} = \underset{d\times k}{A}\;\underset{m\times k}{B'} \qquad k << d, m$$

we will see this again in case study 4

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

square

$$\underset{d\times d}{\Sigma} = \underbrace{\underset{d\times k}{\Lambda}\,\Lambda^T}_{\text{sym. + square}} + \Sigma_0$$

$$\Sigma_0 = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix}$$

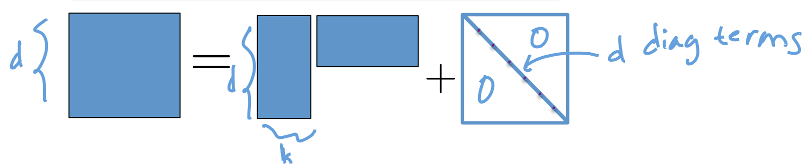
pos. def.

Low-Rank Approximations

- In pictures…

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

$$\underset{d\times d}{\Sigma} = \underset{d\times k}{\Lambda}\Lambda' + \Sigma_0$$



- Number of parameters:

$$dk + d = d(k+1) << \frac{d(d+1)}{2}$$

sig. reduction in param. for $k << d$

Latent Factor Models

here: assume linear term is removed

- Original multivariate regression

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

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

$$y^i = \underset{d\times k}{\Lambda}\,\underset{k\times 1}{\eta^i} + \tilde{\epsilon}^i \qquad \eta^i \sim N_k(0, I)$$

$$\tilde{\epsilon}^i \sim N_d(0, \Sigma_0)$$

obs. "Factor loadings" "latent factors" diag.

- Proof:

$$\mathrm{cov}(y, \Lambda, \Sigma_0) = E[(y - \cancelto{0}{E[y]})(y - \cancelto{0}{E[y]})^T] = E[yy^T]$$

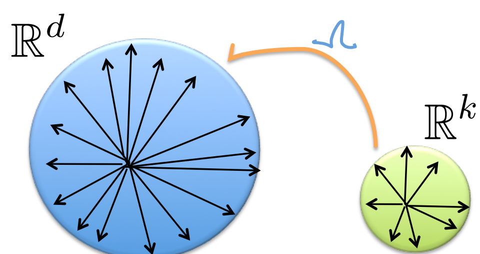
$$= E[(\Lambda\eta + \tilde{\epsilon})(\Lambda\eta + \tilde{\epsilon})^T] = \Lambda E[\eta\eta^T]\Lambda \;+ \cancel{2E[\eta^T]\Lambda^T E[\tilde{\epsilon}]}\; + E[\tilde{\epsilon}\tilde{\epsilon}^T] \rightarrow \Sigma_0$$

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

Lower-dim Embeddings

Sharing information in
low-dim subspace

obs. y^i



η^i latent factor

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