5.4 Maximum A Posteriori Estimator and Lasso Estimator.

In this section, we study the Beyersian model:

$$\eta = X^T\beta + \epsilon$$

where $X \in \mathbb{R}^{p\times n}$

$\beta = (\beta_1, \cdots, \beta_p)^T$ with i.i.d $\beta_i \sim L(0, b)$

$\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$ with i.i.d $\epsilon_i \sim N(0, \sigma_\epsilon^2)$

β and ϵ are independent.

Prop: $\beta^{map} = \beta^{lasso}$ with regularization parameter $\lambda = \frac{2\sigma_\epsilon^2}{b}$.

Proof: The PDFs of β and ϵ are

$$p_\beta(t) = \frac{1}{(2b)^p} e^{-\frac{\|t\|_{\ell^1}}{b}}, \quad p_\epsilon(r) = \frac{1}{\sqrt{(2\pi)^n \sigma_\epsilon^n}} e^{-\frac{\|r\|^2}{2\sigma_\epsilon^2}}$$

To compute $p_{\eta|\beta}(s|t)$, notice that if $\beta = t$ is realized, then $\eta = X^T\beta + \epsilon = \underbrace{X^T t}_{\text{deterministic}} + \underbrace{\epsilon}_{\text{random}}$

HW 10 $\sim N(X^T t, \sigma_\epsilon^2)$, thus

$$p_{\eta|\beta}(s|t) = \frac{1}{\sqrt{(2\pi)^n \sigma_\epsilon^n}} e^{-\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2}}.$$

Take "ln" in the Bayes' Thm $p_{\beta|\eta}(t|s) = \frac{p_{\eta|\beta}(s|t)\, p_\beta(t)}{p_\eta(s)}$:

$$\ln p_{\beta|\eta}(t|s) = \ln p_{\eta|\beta}(s|t) + \ln p_\beta(t) - \ln p_\eta(s)$$

$$= \underbrace{\ln \frac{1}{\sqrt{(2\pi)^n \sigma_\epsilon^n}}}_{\text{const in } t} - \underbrace{\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2}}_{\text{quadratic in } t}$$

$$+ \underbrace{\ln \frac{1}{(2b)^p}}_{\text{const in } t} - \underbrace{\frac{\|t\|_{\ell^1}}{b}}_{\ell^1 \text{ norm of } t}$$

$$-\underbrace{\ln p_\eta(s)}_{\text{const in } t}$$

$$= -\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2} - \frac{\|t\|_{\ell^1}}{b} + \text{const in } t.$$

$$= -\frac{1}{2\sigma_\epsilon^2}\left[\|s - X^T t\|^2 + \frac{2\sigma_\epsilon^2}{b}\|t\|_{\ell^1}\right] + \text{const in } t.$$

Therefore,

$$\beta^{map} \overset{def}{=} \arg\max_t p_{\beta|\eta}(t|s) = \arg\max_t \ln p_{\beta|\eta}(t|s)$$

$$= \arg\min_t \left[\|s - X^T t\|^2 + \frac{2\sigma_\epsilon^2}{b}\|t\|_{\ell^1}\right]$$

$$= \beta^{lasso} \qquad \text{with } \lambda = \frac{2\sigma_\epsilon^2}{b}.$$

∎

Remark: $\beta^{map} = \beta^{lasso}$ (if the realization s is denoted by y)

with i.i.d Laplacian prior.

Summary of this Chapter:

- Basis probability theory and 3 types of linear models with noise.
- Concepts: bias, variance, MSE, β^{blue}, β^{map}
- Conclusions: roughly speaking

$\beta^{blue} \Leftrightarrow \beta^{ls}$

β^{map} with i.i.d Gaussian prior $\Leftrightarrow \beta^{ridge}$

β^{map} with i.i.d Laplacian prior $\Leftrightarrow \beta^{lasso}$

Final Review

Before Midterm: Lecture 17-18

After Midterm:

2 Dimension Reduction:

2.1 kernel PCA: a method to embed sample pts into a high dim space before applying PCA.

Setting: Given sample pts $x^{(1)}, \cdots, x^{(n)} \in \mathbb{R}^p$ and feature map

$$\phi: \mathbb{R}^p \to \mathbb{R}^D \qquad p \ll D$$

$$x^{(i)} \mapsto \phi(x^{(i)}) = [\phi_1(x^{(i)}), \cdots, \phi_D(x^{(i)})]^T$$

The sample matrix in $\mathbb{R}^D$

is $\Phi = [\phi(x^{(1)}), \cdots, \phi(x^{(n)})] \in \mathbb{R}^{D \times n}$

The centered sample matrix is ΦH

(where $H = I - \frac{1}{n} \mathbb{1} \mathbb{1}^T$)

The centered sample covariance matrix

is $\Phi H (\Phi H)^T \in \mathbb{R}^{D \times D}$

Its positive eigenvalues and eigenvectors can

be obtained from the kernel matrix

$$(\Phi H)^T \Phi H = H^T \Phi^T \Phi H$$

Algorithm: Lecture 21

Related Topics: kernel function: $k(x,y) = \phi^T(x)\phi(y)$

kernel matrix: $X^T X$ or $\Phi^T \Phi$

2.2. MDS : a method to find a configuration

of pts in lower dim space that

preserves pairwise dissimilarities.

Setting: given square distance matrix D, the relation between D and the kernel matrix K is

$$D = k \cdot \mathbb{1}^T + \mathbb{1} \cdot k^T - 2K \quad \text{where } k = \text{diag}(K)$$

$$K = -\frac{1}{2} HDH$$

The lower dim space is spanned by the eigenvectors of $-\frac{1}{2} HDH$ associated to the largest two eigenvalues.

Algorithm: Lecture 24

Related Topics:
- more relations between K and D (HW6 #1-4)
- sym. pos semi-def low rank approximation (HW7 #3)