علیرضا درویشی - 96109674 - تمرین ۹ - تحلیل رگرسیون

الف)

$$J(\hat\beta) = [(1-\alpha)\|\beta\|_1 + \alpha\|\beta\|_2^2]$$

$$RSS(\hat\beta) + \lambda J(\hat\beta) = \|Y - X\hat\beta\|_2 + \lambda J(\hat\beta)$$

نشان می دهیم کمینه کردن دو تابع گفته شده، یک جواب می دهد.

فرض کنید $\hat\beta$، کمینه کننده‌ی $RSS+\lambda J$ روی R^{P+1} باشد. حال تعریف می کنیم $t := J(\hat\beta)$

و فرض کنید $\tilde\beta$ کمینه کننده‌ی RSS روی مجموعه‌ی $\{J \le t\}$ است. توجه کنید که $\hat\beta \in \{J \le t\}$

پس:

$$RSS(\tilde\beta) \le RSS(\hat\beta)$$

از طرفی:

$$RSS(\hat\beta) + \lambda J(\hat\beta) \le RSS(\tilde\beta) + \lambda J(\tilde\beta)$$

با ترکیب دو نتیجه داریم:

$$RSS(\hat\beta) + \lambda J(\hat\beta) \le RSS(\tilde\beta) + \lambda J(\tilde\beta) \le RSS(\hat\beta) + \lambda J(\tilde\beta) \le RSS(\hat\beta) + \lambda t = RSS(\hat\beta) + \lambda J(\hat\beta)$$

پس تمام نامساوی ها، تساوی است.

پس $RSS(\hat\beta) + \lambda J(\hat\beta) = RSS(\tilde\beta) + \lambda J(\tilde\beta)$. با فرض محدب بودن $RSS + \lambda J$،

کمینه یکتاست پس $\hat\beta = \tilde\beta$، پس هر دو سوال، نتیجه‌ی مشابه دارند.

از طرفی،

$$\frac{\lambda_1}{\lambda_2} = \frac{1-\alpha}{\alpha}, \quad \min\left(RSS(\beta) + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\right) = t$$

Scanned with CamScanner

ب ،

$$\pi(B) = A\, e^{-K\left(\alpha \sum_{i\geq 1} B_i^2 + (1-\alpha)\sum_{i\geq 1} |B_i|\right)}$$

در این صورت ،

$$\log \pi(B) = Const - K\left(\alpha \sum_{i\geq 1} B_i^2 + (1-\alpha)\sum_{i\geq 1} |B_i|\right)$$

که در این صورت :

$$\hat{\beta} = \underset{\beta}{\arg\max}\left(\log(\pi(\beta)) + \log f(y|\beta)\right)$$

$$= \underset{\beta}{\arg\max}\left(-\frac{1}{\sigma^2} RSS(\beta) - K J(\beta)\right) = \boxed{\underset{\beta}{\arg\min}\left(RSS(\beta) + \lambda J(\beta)\right)}$$

Scanned with CamScanner

PCA

برای بردار اول: $Z_1 = x\phi_1$

هدف این است که $Var(Z_1)$ بیشینه شود پس:

$Z_1 = \sum_i x_i \phi_{1i}$

$$Var(Z_1) = \sum \phi_{1i}\phi_{1j} Cov(x_i, x_j) = \phi_1^T \Sigma \phi_1 , \quad \Sigma_{ij} = Cov(x_i, x_j)$$

حال:

$$\phi_1 = \underset{\phi_1}{\arg\max} (\phi_1^T \Sigma \phi_1), \quad \|\phi_1\| \le 1$$

از طرفی از جبر خطی می‌دانیم اگر Σ یک ماتریس متقارن باشد، ϕ_1 ماکسیمم کننده‌ی بالا، بردار ویژه‌ی متناظر با بیشینه‌ی مقدار ویژه‌های ماتریس Σ است.

حال برای ϕ_k:

$$\phi_k = \underset{\phi}{\arg\max} (\phi^T \Sigma \phi), \quad \|\phi\| \le 1, \quad \phi^T\phi_j = 0 \text{ for } j \in 0, \dots, k-1$$

ϕ_k باید بیشینه‌ترین واریانس باشد، در عین حال بر $\phi_1 \dots \phi_{k-1}$ عمود باشد.

از جبر خطی هم می‌دانیم که ϕ_k، متناظر با بردار ویژه‌ی k امین بزرگترین مقدار ویژه‌ی Σ است.

ب)

در قسمت قبل تعریف شد!

Scanned with CamScanner