ECBM4040 Fall 2016: Homework #2

Pan Chang - cp2923@columbia.edu
Discussants: none

October 28, 2016

Problem 1

First we should notice the functions we deal with are all quadratic, thus convex.

1. We can take derivation directly.

$$\begin{aligned}\frac{\partial \boldsymbol{L}_{ls}}{\partial \boldsymbol{A}} &= \frac{\partial \Sigma_{i=1}^{m}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)^T(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)}{\partial \boldsymbol{A}} \\ &= \frac{\partial \Sigma_{i=1}^{m}(\boldsymbol{y}^{iT}\boldsymbol{y}^i - 2\boldsymbol{y}^{iT}\boldsymbol{A}\boldsymbol{x}^i + (\boldsymbol{A}\boldsymbol{x}^i)^T\boldsymbol{A}\boldsymbol{x}^i)}{\partial \boldsymbol{A}} \\ &= -2\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT} + 2\Sigma_{i=1}^{m}\boldsymbol{A}\boldsymbol{x}^i\boldsymbol{x}^{iT}\end{aligned}$$

Set it to zero then we have:

$$\boldsymbol{A}_{ls} = (\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT})(\Sigma_{i=1}^{m}\boldsymbol{x}^i\boldsymbol{x}^{iT})^{-1}$$

2. Similarly, we have

$$\begin{aligned}\frac{\partial \boldsymbol{L}_{r}}{\partial \boldsymbol{A}} &= \frac{\partial \lambda Tr(\boldsymbol{A}^T\boldsymbol{A})}{\partial \boldsymbol{A}} - 2\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT} + 2\Sigma_{i=1}^{m}\boldsymbol{A}\boldsymbol{x}^i\boldsymbol{x}^{iT} \\ &= 2\lambda A - 2\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT} + 2\Sigma_{i=1}^{m}\boldsymbol{A}\boldsymbol{x}^i\boldsymbol{x}^{iT}\end{aligned}$$

By setting it to zero, we have:

$$\boldsymbol{A}_{lr} = (\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT})(\lambda\boldsymbol{I} + \Sigma_{i=1}^{m}\boldsymbol{x}^i\boldsymbol{x}^{iT})^{-1}$$

3. The likelihood function is:

$$L(\boldsymbol{A}) \propto \Pi_{i=1}^{m}\exp\{-\frac{1}{2}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)^T\sigma^{-2}\boldsymbol{I}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)\}$$

After taking ln, we have:

$$\begin{aligned}l(A) &\propto -\frac{1}{2}\Sigma_{i=1}^{m}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)^T\sigma^{-2}\boldsymbol{I}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i) \\ &\propto \Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT} - \Sigma_{i=1}^{m}\boldsymbol{A}\boldsymbol{x}^i\boldsymbol{x}^{iT}\end{aligned}$$

By taking derivation, we have:

$$\boldsymbol{A}_{ML} = (\Sigma_{i=1}^{m}\boldsymbol{y}^i\boldsymbol{x}^{iT})(\Sigma_{i=1}^{m}\boldsymbol{x}^i\boldsymbol{x}^{iT})^{-1}$$

4. The prior distribution is:

$$f(A) \propto \exp\{-\frac{1}{2}Tr[\lambda(\boldsymbol{A}-\boldsymbol{M})^T(\boldsymbol{A}-\boldsymbol{M})]\}$$

This the posterior distribution is:

$$f(\boldsymbol{A}|\boldsymbol{X}) \propto \exp\{-\frac{1}{2}Tr[\lambda(\boldsymbol{A}-\boldsymbol{M})^T(\boldsymbol{A}-\boldsymbol{M})] - \frac{1}{2}\Sigma_{i=1}^m(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)^T\sigma^{-2}\boldsymbol{I}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)\}$$

Since we only care about MAP, we can take ln on both side.

$$\ln f(\boldsymbol{A}|\boldsymbol{X}) \propto -Tr[\lambda(\boldsymbol{A}-\boldsymbol{M})^T(\boldsymbol{A}-\boldsymbol{M})] - \Sigma_{i=1}^m(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)^T\sigma^{-2}\boldsymbol{I}(\boldsymbol{y}^i - \boldsymbol{A}\boldsymbol{x}^i)$$

After set the derivation to zero, we have:

$$\begin{aligned}\boldsymbol{0} &= -2\lambda(\boldsymbol{A}-\boldsymbol{M}) + 2\sigma^{-2}\Sigma_{i=1}^m \boldsymbol{y}^i\boldsymbol{x}^{iT} - 2\Sigma_{i=1}^m \boldsymbol{A}\boldsymbol{x}^i\boldsymbol{x}^{iT} \\ \Rightarrow \boldsymbol{A}_{MAP} &= (\sigma^{-2}\Sigma_{i=1}^m \boldsymbol{y}^i\boldsymbol{x}^{iT} + \lambda\boldsymbol{M})(\lambda\boldsymbol{I} + \sigma^2\Sigma_{i=1}^m \boldsymbol{x}^i\boldsymbol{x}^{iT})^{-1}\end{aligned}$$

If $\boldsymbol{M} = \boldsymbol{0}$,

$$\boldsymbol{A}_{MAP} = (\sigma^{-2}\Sigma_{i=1}^m \boldsymbol{y}^i\boldsymbol{x}^{iT})(\lambda\boldsymbol{I} + \sigma^2\Sigma_{i=1}^m \boldsymbol{x}^i\boldsymbol{x}^{iT})^{-1}$$

5. We can see that the estimator given by least square and maximum likelihood are the same if the distribution of ϵ_i is normal.

The estimator given by least square with a Frobenius penalty and maximum a posterior are the same if $\sigma = 1$ and $\boldsymbol{M} = \boldsymbol{0}$.