Exercise 5

Bowen Hua

October 10, 2017

1 Penalized likelihood and soft thresholding

1.1 (A)

1.1.1 Our objective above is the negative log likelihood of a Gaussian distribution.

Suppose $y_i \sim \mathcal{N}(\theta, 1)$, then the negative log likelihood function is

$$-\log(\frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\theta)^2}{2}}) = -\log(\frac{1}{\sqrt{2\pi}}) + (y-\theta)^2/2 = -\log(\frac{1}{\sqrt{2\pi}}) + \frac{1}{2}(y-\theta)^2$$

Ignoring the constant term, we have our desired result.

1.1.2 Derive the minimizer

We reformulate the problem as a constrained optimization problem:

$$\begin{aligned} \min_{\theta,z} \quad & \frac{1}{2}(y-\theta)^2 + \lambda z && && (1) \\ \text{s.t.} \quad & z \geq \theta && : \mu_1 && (2) \\ & z \geq -\theta && : \mu_2 && (3) \end{aligned}$$

The dual variables are displayed to the right of their corresponding constraints.

The Lagrangian of this optimization problem is:

$$\mathcal{L}(\theta, z, \mu_1, \mu_2) = \frac{1}{2}(y-\theta)^2 + \lambda z + \mu_1^T(\theta - z) + \mu_2^T(-\theta - z)$$

Since the problem is convex, its KKT conditions are necessary and sufficient:

$$\begin{aligned} & \theta - y + \mu_1 - \mu_2 = \mathbf{0} && (4) \\ & \lambda\mathbf{1} - \mu_1 - \mu_2 = \mathbf{0} && (5) \\ & z \geq \theta && (6) \\ & z \geq -\theta && (7) \\ & \mu_1 \geq \mathbf{0} && (8) \\ & \mu_2 \geq \mathbf{0} && (9) \\ & \mu_1 \cdot (\theta - z) = \mathbf{0} && (10) \\ & \mu_2 \cdot (-\theta - z) = \mathbf{0} && (11) \end{aligned}$$

where the operator $\cdot$ denotes element-wise multiplication.

We can see from the KKT conditions that the problem is **separable across each component of** θ. Therefore, we can focus on one component θ_i. Since $\lambda = \mu_{1i} + \mu_{2i}$, we have three cases:

1. $\mu_{1i} = \lambda, \mu_{2i} = 0$. In this case, we have $\theta_i = z_i$ because of complementary slackness. As a result, $\theta_i \geq 0$. Plugging this information into (4), we get $\theta_i = y_i - \lambda$. We also know that in this case $y_i \geq \lambda$.

2. $\mu_{2i} = \lambda, \mu_{1i} = 0$. In this case, we have $\theta_i = -z_i$ because of complementary slackness. As a result, $\theta_i \leq 0$. Plugging this information into (4), we get $\theta_i = y_i + \lambda$. We also know that in this case $y_i \leq -\lambda$.

3. $\lambda = \mu_{1i} + \mu_{2i}, \mu_{1i} > 0, \mu_{2i} > 0$. In this case, we have dual degeneracy. Complementary slackness implies $\theta_i = z_i = 0$. We also know that in this case $-\lambda < y_i < \lambda$.

Putting the three cases together, we can see that we have the desired result.