
Solution 1

(a) Writing $\widehat{\mu} - \mu = H_\lambda y - \mu = H_\lambda(y-\mu) + (H_\lambda - I)\mu$ gives

$$(\widehat{\mu}-\mu)^{\mathrm{T}}(\widehat{\mu}-\mu) = (y-\mu)^{\mathrm{T}} H_\lambda^{\mathrm{T}} H_\lambda (y-\mu) + 2(y-\mu)^{\mathrm{T}} H_\lambda^{\mathrm{T}}(H_\lambda - I)\mu + \mu^{\mathrm{T}}(H_\lambda - I)^{\mathrm{T}}(H_\lambda - I)\mu.$$

The expected value of the second term is zero, because $\mathrm{E}(y-\mu)=0$, and the final term is the constant $\|(I-H_\lambda)\mu\|_2^2$, while the first term has expectation

$$\mathrm{E}\left[\mathrm{tr}\left\{(y-\mu)^{\mathrm{T}} H_\lambda^{\mathrm{T}} H_\lambda (y-\mu)\right\}\right] = \mathrm{tr}\left[\mathrm{E}\left\{(y-\mu)(y-\mu)^{\mathrm{T}} H_\lambda^{\mathrm{T}} H_\lambda\right\}\right] = \mathrm{tr}\left(\sigma^2 I_n H_\lambda^{\mathrm{T}} H_\lambda\right),$$

which gives the result.

(b) As $\mathrm{E}(yy^{\mathrm{T}}) = \mathrm{var}(y) + \mathrm{E}(y)\mathrm{E}(y)^{\mathrm{T}}$ we have

$$\begin{aligned}
\mathrm{E}\left[(y-\widehat{\mu})^{\mathrm{T}}(y-\widehat{\mu})\right] &= \mathrm{E}\left[y^{\mathrm{T}}(I-H_\lambda)^{\mathrm{T}}(I-H_\lambda)y\right] \\
&= \mathrm{E}\left[\mathrm{tr}\left\{(I-H_\lambda)^{\mathrm{T}}(I-H_\lambda)yy^{\mathrm{T}}\right\}\right] \\
&= \mathrm{tr}\left\{(I-H_\lambda)^{\mathrm{T}}(I-H_\lambda)(\mu\mu^{\mathrm{T}} + \sigma^2 I_n)\right\} \\
&= \mu^{\mathrm{T}}(I-H_\lambda)^{\mathrm{T}}(I-H_\lambda)\mu + \sigma^2(n - 2\nu_1 + \nu_2) \\
&= \|(I-H_\lambda)\mu\|_2^2 + \sigma^2(n-2\nu_1+\nu_2),
\end{aligned}$$

so $\widehat{\sigma}_\lambda^2$ is biased upwards unless $\|(I-H_\lambda)\mu\|_2 = 0$, i.e., unless μ lies in the kernel of $I - H_\lambda$. In a standard setting H_λ is a projection matrix, so $H_\lambda^{\mathrm{T}} H_\lambda = H_\lambda H_\lambda = H_\lambda$, and thus $\nu_1 = \nu_2 = \mathrm{rank}(H_\lambda)$, so $\widehat{\sigma}_\lambda^2$ becomes the usual unbiased variance estimator for a linear model.

Solution 2

(a) We have $X^{\mathrm{T}}X = I_p$ and hence $\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y = X^{\mathrm{T}}y$. The sum of squares is

$$(y - X\beta)^{\mathrm{T}}(y-X\beta) = y^{\mathrm{T}}y - 2y^{\mathrm{T}}X\beta + \beta^{\mathrm{T}}X^{\mathrm{T}}X\beta = y^{\mathrm{T}}y - 2\widehat{\beta}^{\mathrm{T}}\beta + \beta^{\mathrm{T}}\beta,$$

so the function to be minimised is

$$L = \tfrac{1}{2}\left(y^{\mathrm{T}}y - 2\widehat{\beta}^{\mathrm{T}}\beta + \beta^{\mathrm{T}}\beta\right) + \lambda\sum_{r=1}^{p}|\beta_r| \equiv \sum_{r=1}^{p}(\beta_r^2/2 - \widehat{\beta}_r\beta_r + \lambda|\beta_r|).$$

This is a sum of p separate functions, each of which is a sum of the two convex functions of the forms $x^2 - ax$ and $b|x|$ for $b > 0$, so each is convex. Each of the p individual summands can be minimised individually.

(b) The function L is differentiable in β_r except at $\beta_r = 0$, and the minimum is either at $\beta_r = 0$ or elsewhere. Now

$$\partial L/\partial\beta_r = \beta_r - \widehat{\beta}_r + \lambda\,\mathrm{sign}(\beta_r),$$

so

$$\lim_{\beta_r\to 0_+}\partial L/\partial\beta_r = \lambda - \widehat{\beta}_r, \quad \lim_{\beta_r\to 0_-}\partial L/\partial\beta_r = -\lambda - \widehat{\beta}_r.$$

- For a minimum at $\beta_r = 0$ we must have $\lambda - \widehat{\beta}_r > 0$ and $-\lambda - \widehat{\beta}_r < 0$, or equivalently $|\widehat{\beta}_r| < \lambda$. In this case $\tilde{\beta}_r = 0$.
- For a minimum not at $\beta_r = 0$, setting $\partial L / \partial \beta_r = 0$ gives

  $$\tilde{\beta}_r = \widehat{\beta}_r - \lambda \,\mathrm{sign}(\tilde{\beta}_r),$$

  so if $\tilde{\beta}_r > 0$, then $\tilde{\beta}_r = \widehat{\beta}_r - \lambda$, whereas if $\tilde{\beta}_r < 0$, then $\tilde{\beta}_r = \widehat{\beta} + \lambda$.

Putting these two cases together gives

$$\tilde{\beta}_r = \mathrm{sign}(\widehat{\beta}_r)(|\widehat{\beta}_r| - \lambda)I(|\widehat{\beta}_r| > \lambda),$$

as required.