Stat/CS 5525 Homework 2

Kazi Hasan Ibn Arif / hasanarif@vt.edu / SID: 906614469

September 25, 2023

Question 2

We aim to show that the ridge regression estimate is both the mean and the mode of the posterior distribution for β, under a Gaussian prior $\beta \sim \mathcal{N}(0, \tau^2 I)$ and a Gaussian likelihood $y \sim \mathcal{N}(X\beta, \sigma^2 I)$.

The ridge regression estimate is given by:

$$\hat{\beta} = \left(X^T X + \lambda I\right)^{-1} X^T Y$$

We assume a Gaussian prior for β:

$$\beta \sim \mathcal{N}(0, \tau^2 I)$$

and a Gaussian likelihood for the data y:

$$y|\beta \sim \mathcal{N}(X\beta, \sigma^2 I)$$

The posterior distribution $Pr(\beta|y, X)$ is proportional to the product of the likelihood and the prior:

$$Pr(\beta|y, X) = \frac{1}{K} Pr(y|\beta, X) Pr(\beta)$$

where $K = \int Pr(y|\beta, X) Pr(\beta) d\beta$ is independent of β.

By substituting the Gaussian forms for the likelihood and prior, we get:

$$\log(Pr(\beta|y, X)) = -C - \frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2} - \frac{\beta^T \beta}{2\tau^2}$$

where C collects all terms that do not depend on β.

This expression is maximized when $\hat{\beta} = \left(X^T X + \frac{\sigma^2}{\tau^2} I\right)^{-1} X^T y$. Setting $\lambda = \frac{\sigma^2}{\tau^2}$, we find that this is equivalent to the ridge regression estimate.

It is straightforward to see that the posterior distribution is Gaussian, implying that its mean and mode coincide. Let $m = \hat{\beta}$. The inverse of the covariance matrix Σ satisfies:

$$\Sigma^{-1} = \frac{1}{\sigma^2} \left(X^T X + \frac{\sigma^2}{\tau^2} I\right)$$

Thus $\hat{\beta} = \frac{1}{\sigma^2}\Sigma X^T y$, confirming that $\hat{\beta}$ must be the mean of the posterior distribution.

The ridge regression estimate is both the mean and the mode of the posterior distribution under the Gaussian prior and likelihood. The regularization parameter λ is related to τ^2 and σ^2 by $\lambda = \frac{\sigma^2}{\tau^2}$.