# MLE 5

Sonntag, 18. Juni 2023 11:07

## 1 Bias and variance of ridge regression (8 points)

Ridge regression solves the regularized least squares problem

$$\widehat{\beta}_\tau = \mathrm{argmin}_\beta (y - X\beta)^\top (y - X\beta) + \tau\, \beta^\top \beta$$

with regularization parameter $\tau \geq 0$. Regularization introduces some bias into the solution in order to achieve a potentially large gain in variance. Assume that the true model is $y = X\beta^* + \epsilon$ with zero-mean Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and centered features $\frac{1}{N}\sum_i X_i = 0$ (note that these assumptions imply that $y$ is also centered in expectation). Prove (e.g. using the SVD of $X$) that expectation and covariance matrix of the regularized solution (both taken over all possible training sets of size $N$) are then given by

$$\mathbb{E}[\widehat{\beta}_\tau] = S_\tau^{-1} S\, \beta^* \qquad \mathrm{Cov}[\widehat{\beta}_\tau] = S_\tau^{-1} S\, S_\tau^{-1} \sigma^2$$

where $S$ and $S_\tau$ are the ordinary and regularized scatter matrices:

$$S = X^\top X \qquad S_\tau = X^\top X + \tau \mathbb{I}_D$$

Notice that expectation and covariance reduce to the corresponding expressions of ordinary least squares (as derived in the lecture) when $\tau = 0$:

$$\mathbb{E}[\widehat{\beta}_{\tau=0}] = \beta^* \qquad \mathrm{Cov}[\widehat{\beta}_{\tau=0}] = S^{-1}\, \sigma^2$$

Since $S_\tau$ is greater than $S$ (in any norm), regularization has a shrinking effect on both expectation and covariance.

$$\mathbb{E}[\hat{\beta}_T] = (X^TX + \tau \mathbb{I}_D)^{-1} X^T y$$
$$= (X^TX + \tau \mathbb{I}_D)^{-1} X^T (X\beta^* + \varepsilon)$$
$$= \underbrace{(X^TX + \tau \mathbb{I}_D)^{-1}}_{S_T^{-1}} (\underbrace{X^TX}_{S}\beta^* + \underbrace{X^T\varepsilon}_{\mathbb{E}=0})$$
$$\Rightarrow \mathbb{E}(\hat{\beta}_T) = S_T^{-1} S \beta^*$$

$$\text{bias} = \hat{\beta}_\tau - \beta^* = (X^TX + \tau \mathbb{I}_D)^{-1} \cdot (X^TX\beta^* + X^T\varepsilon) - \beta^*$$
$$\mathrm{cov}(\hat{\beta}_T) = \mathbb{E}\big((\hat{\beta}_\tau - \underbrace{\mathbb{E}(\hat{\beta}_\tau)}_{\beta^*})(\hat{\beta}_T - \underbrace{\mathbb{E}(\hat{\beta}_T)}_{\beta^*})^T\big)$$
$$= \mathbb{E}\Big[\big((X^TX + \tau \mathbb{I}_D)^{-1}(X^TX\beta^* + X^T\varepsilon) - \beta^*\big)\big((\beta^{*T}XX^T + \varepsilon^T X)(X^TX + \tau \mathbb{I}_D)^{-1} - \beta^{*T}\big)\Big]$$

↯ Most Probably, there is something wrong here which I can't seem to find / solve

**Number 2:**

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y_i^* - X_i \cdot \beta)^2 \overset{!}{=} 0 \qquad (1)$$

$$\Sigma \cdot \beta + \frac{1}{4}(\mu_1 - \mu_{-1})^T \cdot (\mu_1 - \mu_{-1}) \cdot \beta = \frac{1}{2}(\mu_1 - \mu_{-1})^T \qquad (2)$$

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{N} \left(y_i^{*2} - 2y_i^* X_i \beta + (X_i\beta)^2\right) = 0$$
$$\Rightarrow \sum_{i=1}^{N} \left(-2y_i^* X_i + 2X_i^2 \beta\right) = 0$$
$$\Rightarrow 2 \cdot \sum_{i=1}^{N} \left(X_i^2 \beta - y_i^* X_i\right) = 0$$
$$\Rightarrow \sum_{i=1}^{N} (X_i^2 \beta) = \sum_{i=1}^{N} y_i^* X_i$$
$$\Rightarrow \beta \cdot \sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} y_i^* X_i$$
$$\Rightarrow \Sigma \beta \sum_{i=1}^{N} X_i^2 = \Sigma \sum_{i=1}^{N} y_i^* X_i \quad ?$$

I have no clue how to go on from this or if it is even remotely correct so far