Lab 14

Machine Learning 2021-2022 - UMONS
Souhaib Ben Taieb

1

Consider the problem of multiple linear regression, where the aim is to find $\hat{\beta^{LS}} \in \mathbb{R}^p$ such that:

$$\hat{\beta}^{LS} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2$$

We can show that the ordinary least squares estimate is given by $\hat{\beta}^{LS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{y}$ where $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\boldsymbol{y} \in \mathbb{R}^n$. Now consider the problem of ridge regression, where the optimization problem is now formulated as finding $\hat{\beta}^R \in \mathbb{R}^p$ such that:

$$\hat{\beta}^{R} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

The solution is given by $\hat{\beta}^R = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}_p)^{-1} \mathbf{X}^\top \boldsymbol{y}$, where $\mathbf{I}_p$ is the $p \times p$ identity matrix. Assuming that $p = n$ and that $\mathbf{X} = \mathbf{I}_p$,

- Prove that $\hat{\beta}^R = \frac{\hat{\beta}^{LS}}{\lambda + 1}$.
- Assuming that $\text{Bias}(\hat{\beta}^{LS}) = 0$, compute the bias of the ridge estimator.
- Derive the covariance matrices of $\hat{\beta}^{LS}$ and $\hat{\beta}^R$ and show how they relate to one another. Assume that $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \mathbf{I}_n$, where $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is a random noise vector (i.e. $\boldsymbol{y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$).

2

Suppose that the columns of $\mathbf{X}_1$ are orthonormal, and that $\mathbf{X}_2 = 10\mathbf{X}_1$. Show that the ordinary least squares estimates are equivariant, meaning that multiplying $\mathbf{X}$ by a constant c scales the coefficients estimates by a factor $\frac{1}{c}$.

Is it also the case for the ridge estimates ?