Machine learning in empirical economic research

Problem set

Andreas Dzemski

May 29, 2019

Problem 1

The Ridge regressor is given by

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}^{p_n}} \sum_{i=1}^{n} \left(y - \beta_0 - x_i' \beta_1\right)^2 + \lambda \|\beta_1\|_2^2$$

1. Show that Ridge regression minimizes a strictly convex function and conclude that $\hat{\beta}^{\text{ridge}}$ is always uniquely defined.

2. Show that

$$\hat{\beta}^{\text{ridge}} = \left(\mathbf{X}'\mathbf{X} + \lambda \operatorname{diag}\left((0, 1, \dots 1)'\right)\right)^{-1} \mathbf{X}'\mathbf{y}$$

3. Suppose that $p_n < n$ and that the off-diagonal elements of $\mathbf{X}'\mathbf{X}/n$ are zero. In that case $\hat{\beta}^{\text{ols}}$ is defined. Show that

$$\left(\frac{(\mathbf{X}'\mathbf{X}/n)_{jj}}{(\mathbf{X}'\mathbf{X}/n)_{jj} + \lambda}\right) \hat{\beta}^{\text{ols}}.$$

   Relate this result to the "shrinkage" property of Ridge regression.

Problem 2

The L_q-penalized least squares estimator is given by

$$\hat{\beta}^{L_q} = \arg\min_{\beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}^{p_n}} \sum_{i=1}^{n} \left(y - \beta_0 - x_i' \beta_1\right)^2 + \lambda \|\beta_1\|_q^q.$$

1. Show that this definition is equivalent to

$$\hat{\beta}^{L_q} = \arg\min_{\beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}^{p_n}} \sum_{i=1}^{n} \left(y - \beta_0 - x_i'\beta_1\right)^2$$
$$\text{subject to:} \quad \|\beta_1\|_q \leq s_\lambda$$

   for some s_λ.

2. Verify for L_q-penalized linear regression that $\mathcal{F}_\lambda \subset \mathcal{F}_{\lambda'}$ for $\lambda < \lambda'$.

3. Let $\mathbf{X}_{+1}$ denote the design matrix including intercept ($n \times (p_n+1)$ matrix). Suppose that $\mathbf{X}_{+1}'\mathbf{X}_{+1}/n$ has full rank. Let $\hat{\beta}^{\text{ols}}$ denote the OLS estimator

$$\hat{\beta}^{\text{ols}} = \left(\mathbf{X}_{+1}'\mathbf{X}_{+1}\right)^{-1} \mathbf{X}_{+1}'\mathbf{y}.$$

   Show that the contour sets

$$\left\{\beta \in \mathbb{R}^{p_n+1} : \|\mathbf{y} - \mathbf{X}_{+1}\beta\|_2^2 = c\right\}$$

   indexed by $c \in \mathbb{R}$ are empty or ellipsoids centered at $\hat{\beta}^{\text{ols}}$.

Problem 3

We consider the the simulation excercise from the slides for the “sparse uncorrelated” design. The repository with the simulation code used to generate the slides can be found at `https://github.com/adzemski/ML_notes`.

1. In the Gaussian case $\hat{\beta}_1^{\text{ols}}$ is the MLE estimator. Explain the efficiency result for MLE estimators (Cramér-Rao lower bound). Theoretically, can $\hat{\beta}_1^{\text{post}}$ beat $\hat{\beta}_1^{\text{ols}}$ in terms of efficiency?

2. Extend the simulation excercise to simulate also the variance and the MSE of $\hat{\beta}_1^{\text{post}}$ and $\hat{\beta}_1^{\text{ols}}$. Which estimator is more efficient?

Problem 4

Implement and simulate the double selection procedure from Belloni, Chernozhukov, and Hansen (2014) for the “sparse correlated” desgin from the slides. The repository with the simulation code used to generate the slides can be found at `https://github.com/adzemski/ML_notes`.