STATISTICAL METHODS IN FINANCE, PROBLEM SHEET 7
MSC IN MATHEMATICS AND FINANCE, 2023-2024

CRIS SALVI, WILL TURNER

Exercise 1 (Multiple choice)**.** Answer the following questions by Yes/No/Maybe, and justify your answer.

(i) In a simple (one-dimensional) linear regression,
   (a) if the value of R^2 is equal to one, is the cloud of points along a perfect line?
   (b) does the optimal least-square line go through the point $(\overline{x}, \overline{y})$?
   (c) is the vector $\widehat{\mathbf{Y}}$ always orthogonal to the estimated residuals $\widehat{\boldsymbol{\varepsilon}}$?

(ii) Are the two estimators $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{ML}}$, respectively obtained by Least squares and by Maximum likelihood, the same?

Exercise 2. In this problem, we would like to study whether the variations of the COCA-COLA (ticker: KO) returns can be explained by those of PEPSI (ticker: PEP). We consider the two-year period over 2016 and 2017. Write an IPython notebook performing the following tasks:

- a linear regression of KO (output variable $\mathbf{Y}$) vs PEP (input variable $\mathbf{X}$) of the form $\mathbf{Y} = \beta\mathbf{X} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.
- Compute the empirical averages $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$.
- Compute the analytical value of the optimal least-square estimator $\widehat{\boldsymbol{\beta}}$ and compare it to the output of the regression.
- Compute the estimator of σ^2.
- Test the hypothesis $\beta = 0$ with a 95% confidence interval.
- Assuming the PEPSI returns are Gaussian over 2018, construct a day-by-day forecast of the KO returns.

Exercise 3 (Generalized Ridge Regression)**.** Consider a problem of Generalized Ridge Regression, where one seeks to find $\widehat{\boldsymbol{\beta}}$ that minimizes

$$\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{P}} + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2_{\mathbf{Q}} \tag{0.1}$$

where $\|\boldsymbol{\beta}\|^2_{\mathbf{P}} = \boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}$ stands for the weighted norm squared, $\mathbf{P}$ and $\mathbf{Q}$ are positive definite matrices and vector $\boldsymbol{\beta}_0$ is the ridge target. Note that for $\mathbf{P} = \mathbf{Q} = \mathbf{I}$ and $\boldsymbol{\beta}_0 = 0$ we obtain the ordinary ridge regression. Derive the generalized ridge coefficient $\widehat{\boldsymbol{\beta}}$ that minimizes (0.1), and argue about its uniqueness.

Hint: Use the fact $\mathbf{P}$ and $\mathbf{Q}$ are positive definite when trying to show a certain matrix is invertible.

Date: May 9, 2024.

Exercise 4 (Robust view of LASSO)**.** Consider a robust version of the standard linear regression problem, in which we wish to protect ourselves against perturbations of the explanatory variables. To this end we consider

$$\min_{\boldsymbol{\beta}} \max_{\Delta \in \mathcal{U}} \{\|\mathbf{y} - (\mathbf{X} + \boldsymbol{\Delta})\boldsymbol{\beta}\|_2\},$$

where the allowable perturbations $\boldsymbol{\Delta} := (\boldsymbol{\delta}_1, \ldots \boldsymbol{\delta}_p)$ belong to

$$\mathcal{U} := \left\{(\boldsymbol{\delta}_1, \ldots \boldsymbol{\delta}_p) : \left\|\boldsymbol{\delta}_j\right\|_2 \leq c_j \text{ for all } j = 1, \ldots, p\right\} \subset \mathbb{R}^{N \times p}.$$

Hence each feature value x_{ij} can be perturbed by a maximum amount c_j, with the l_2-norm of the overall perturbation vector for that feature bounded by c_j. The perturbations for different features also act independently of one another. We seek the coefficients that minimize squared error under the "worst" allowable perturbation of the features. We also assume that both $\mathbf{y}$ and the columns of $\mathbf{X}$ have been standardized, and have not included an intercept.

Show that the solution to this problem is equivalent to the so-called square-root LASSO

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2 + \sum_{j=1}^{p} c_j \left|\beta_j\right| \right\}.$$

In the special case $c_j = \lambda$ for all $j = 1, 2, \ldots, p$, we thus obtain the square-root LASSO, so that it can be viewed as a method for guarding against uncertainty in the measured predictor values, with more uncertainty leading to a greater amount of shrinkage.

Note that the LASSO is not the same as square-root LASSO, but their objectives are equivalent up to a regularization parameter. So the above equivalence can still provide a robust view of the LASSO.

Exercise 5 (Model selection)**.** The topic of this task is model selection and the bias-variance trade-off. Simulate a dataset of size $n = 100$ consisting of the target variable $\mathbf{y}$ and the independent variables $\mathbf{x}_1, \ldots, \mathbf{x}_{10}$. Both target and explanatory variables are standard normally distributed and independent.

(i) Split the data into two parts: the training data (first 50 data points) and the test data (last 50 data points). Fit the following cascade of models using just the training data:
- the empty model
- $\mathbf{y}$ on $\mathbf{x}_1$
- $\mathbf{y}$ on on $\mathbf{x}_1$ and $\mathbf{x}_2$
- etc.
- full model

For each of the 11 models compute the mean squared error (MSE) on the training set. What do you notice?

(ii) Now compute the MSE for each of the models on the test data and compare the results with the previous part. Which method gives you a better estimate of the expected loss and why?

(iii) Perform shrinkage and variable selection using Ridge and LASSO regressions on the full dataset. Plot the cross-validated MSE curves for Ridge and LASSO regression. Report the minimal cross-validated MSEs and the corresponding values of the penalization parameter λ for Ridge and for LASSO.

(iv) Since you simulated the data independently, the true model is $\mathbf{y} = \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ denotes the error term. Consider your result regarding this. Which method seems to work best/worst and why?

Department of Mathematics, Imperial College London
Email address: wft17@ic.ac.uk