HMW 2: due March 21, 4pm, upload on Canvas

Instructions:

- submit ONE PDF file containing the entire homework; we don't accept multiple files, even when uploaded in a ZIP folder
- include the name of each group member and your student numbers into the PDF
- include in the PDF the code you ran (if you use R Markdown it will be easy to do so).

Question 1

Ward (2017, Journal of Applied Econometrics) forecasts financial crises using a dataset from 1870-2014 on many macroeconomic variables. The goal is to predict the binary variable crisis with the rest of the predictors. The dataset is $\mathrm{R}_c lass$.csv, and its description can be found in Ward (2013) and in the readme file provided on Canvas. In Table 3 in the paper, Ward (2003) contrasts the perfomance of one tree with bagging and a random forest, and the logit benchmark results are in Section 4.1. The R code for this Table is CT_longrun.R; please use this code if you are using R and adapt it to your needs to answer this question *(if you wish to use other software, that is fine, but that means you may have to rewrite the code in another software).*

Just consider the many predictors case in Table 3. Choose from the following alternative methods two: principal component regression (with a logit model in the second step), partial least squares (with a logit model), lasso (with logit) or ridge (with logit). Recreate the results in Table 3 as much as possible using the two methods you chose, and contrast their performance with the results provided in Table 3 by Ward (2017) (if you prefer to report other metrics than the AUC, you are welcome to do so). Also discuss why do you think the results you obtain are better or worse than Ward (2017); I am looking for an answer that also includes intuition about the specific data you are analyzing.

Note: there is no one good answer to this question, and solutions will therefore not be provided; rather, we are looking for a coherent and clear response.

Question 2

a) Write down the objective function for estimating a linear regression model using plain lasso and adaptive lasso. Explain in words when one should use each of these, and how to pick the tuning parameters in each of these regressions.

b) Typically researchers use adaptive lasso as a model selection device: to select covariates. However, they re-estimate the model with the selected covariates in the second step, and this method is called post-adaptive lasso. Explain in words why one would want to report coefficients from a post-adaptive lasso model rather than from the original adaptive lasso procedure.

Question 3

For this question, use the WLLN from the handout on WLLN and CLT (posted under Homework 2). Let the true model be $y_i = X_i'\beta + \epsilon_i, i = 1, \ldots, n$, with y_i a scalar, X_i a $p \times 1$ vector of observed covariates, and β a $p \times 1$ vector of unknown parameters. In matrix form, this model is $y = X\beta + \epsilon$, where y is $n \times 1$ and X is $n \times p$. Let the ridge estimator for $\lambda > 0$ be $\hat{\beta}(\lambda) = (X'X + \lambda I_p)^{-1} X'y$. Below, we hold λ fixed, unless stated otherwise, and p is a fixed number of parameters. You may use a) to solve b); a) and b) to solve c) etc. Assume throughout the exercise that:

Assumption 1 *(i) $X'X$ is invertible; (ii) $E(\epsilon|X) = 0$; (iii) $Var(\epsilon|X) = \sigma^2 I_T$.*

Let $W(\lambda) = (X'X + \lambda I_p)^{-1}$.

a) Show that $E[\hat{\beta}(\lambda)|X] - \beta = -\lambda W(\lambda)\beta$, therefore that the ridge estimator is biased.

b) Show that $Var(\hat{\beta}(\lambda)|X) = W(\lambda)(\sigma^2 X'X)W(\lambda)$ and that $Var(\hat{\beta}(0)|X) - Var(\hat{\beta}(\lambda)|X)$ is positive definite, therefore that the ridge estimator is more efficient than the OLS estimator.

c) Let $PMSE(\lambda|X) = E\|X\hat{\beta}(\lambda) - X\beta\|^2$ (PMSE stands for predictive mean-squared error, a measure related but not equal to the MSE; here $\|\cdot\|$ is the Euclidean norm). Show that

$$PMSE(0|X) - PMSE(\lambda|X) = \lambda \, \text{trace} \, \{W^2(\lambda)(2\sigma^2(X'X) + \lambda(\sigma^2 I_p - \beta\beta' X'X))\}.$$

d) Show that if $\sigma^2 I_p - \beta\beta' X'X$ is positive definite, then the ridge estimator dominates the OLS estimator in predictive mean squared error, in the sense that $PMSE(0|X) - PMSE(\lambda|X) > 0$.

e) Assume that $T^{-1}X'X \xrightarrow{p} \Sigma$, a positive definite matrix of constants, and $\lambda = aT^{\alpha}$, where $a > 0$ is a constant. If $\alpha \in (0, 1/2)$, show that $PMSE(0|X) - PMSE(\lambda|X) \xrightarrow{p} 0$, but if $\alpha = 1/2$, then show that $PMSE(0|X) - PMSE(\lambda|X) \xrightarrow{p} c$, where $c = -a^2(\beta'\Sigma^{-1}\beta) < 0$.