STA6703 SML HW6

Directions

Please submit **one PDF** file including all your reports (answer + code + figures + comments; must be easily readable and have file size under a few megabytes) and **one R or Python code script**. The R/Python script is supplementary material to ensure that your code runs correctly. If you are using RMarkdown, please also include your `.Rmd` file.

Place these two (or three) files in a folder, make a zip or rar archive, and submit the archive electronically via Dropbox file request at `tinyurl.com/nbliznyuk-submit-files` (on the landing page, enter your name so that we know it is you and email so that you get a confirmation).

Deadline: 01-Nov-2022, 10:00 PM EST.

Practice/Optional Problems (do not submit)

1. Complete the R tutorial in ISLR sections 6.5-6.7. You may find the Youtube videos by Trevor Hastie helpful; for links, see file `!_youtube_lab_links.txt` in the subfolder `"[2].code/islr_labs/"`
2. Implementing normal-theory linear regression model with an L2 penalty by hand: extend your implementation of the negative (Gaussian) log-likelihood from hw3 by adding the ridge penalty. Can this objective function be minimized analytically? If the L_2 penalty is replaced by the L_1 penalty, can the resulting objective function be minimized analytically? Briefly explain.
3. Implementing a logistic regression model with an L_2 penalty by hand: extend your implementation of logistic regression (with multiple covariates) from hw4 by adding the ridge penalty. Can this objective function be minimized analytically?
4. "Honest" C_p in the multiple linear regression: the version this criterion motivated by the ISLR authors as the "training MSE corrected for overfitting" is somewhat deficient in assuming either that σ^2 is known or is estimated by $\widehat{\sigma}^2$ (independently of the RSS for each given model fit). Suppose $\widehat{\sigma}^2 = RSS/(n-k-1)$, where the RSS comes from the current model fit; i.e., $\widehat{\sigma}^2$ will be different for different models (even if k is the same). Show that, still, C_p is is an increasing (linear) function of RSS (with the slope and intercept independent of the RSS). Hence, conclude that ranking the models with exactly k predictors with respect to C_p is equivalent to ranking them with respect to RSS.
5. "Honest" AIC and BIC in the multiple linear regression: for the multiple linear regression with iid $Normal(0, \sigma^2)$ errors (the same version considered in class after ch.03), show that the deviance is an increasing function of RSS. Hence conclude that, for a fixed k (hence, fixed d), ranking the models with exactly k predictors using RSS, AIC and BIC produces the same ordering (and hence the best model).

Required Problems (for submission)

ISLR ch.6: 1,2,4,8