1

I. Basic Linear Regression Model.

$\min_{\beta_j} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij} \beta_j)^2$

observed value for dependent variables.

p: # Predictors; X_{ij} : jth predictor for ith observation; β_j: coefficient for jth predictor.

II. LASSO (Least Absolute Shrinkage and Selection Operator):

Adding a penalty term (absolute value of the coefficients) to Basic Linear Regression Model.

$\min_{\beta_j} (\sum_{i=1}^{n} (y_i - \underbrace{\sum_{j=1}^{p} X_{ij} \beta_j}_{\hat{y}_i})^2 + \lambda \sum_{j=1}^{p} |\beta_j|)$ λ: tuning parameter, controls the strength of the penalty. $\lambda \uparrow$, more β_j's are set to 0, leading to a simpler model with fewer variables.

→ Shrinkage: $\lambda \sum |\beta_j|$ term, shrinking coefficients towards 0, preventing overfitting, making the model more interpretable.

→ Selection: when λ is large → setting some coefficients exactly to 0, effectively select variables by excluding some variables from the model entirely.

→ Tuning parameter λ: selected via Cross-validation: different λ's are tried, the one with best predictive performance (lowest (*) cross-validated mean square error) is chosen.

→ LASSO works best (reduce variance of predictions the most.) when the underlying model is "sparse" (many predictors have no effect).

→ LASSO is particularly useful when there are large number of predictors and many of them are irrelevant / redundant. LASSO provides a tradeoff b/w "fitting the data well" & "not being overly complex".

(*) Calculate $MSE = \frac{1}{n}[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|]$ "fold" and computing the avg. of them to get cross-validated mean square error.

Eg. 100 data → 10 folds, each have 10 data. → Use 9 folds to train the model: find most desirable λ, "λ_1"

② then use the 10th fold to test. Repeat this procedure 10 times and avg. $\lambda_1, \lambda_2, \dots, \lambda_{10}$ to get $\bar{\lambda}$ as our final result.

III. Ridge method. (focus only on regularization, no variable selection).

→ $\min_{\beta_j} \left[\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij} \beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$ λ: tuning param that controls penalty strength; $\lambda \sum_{j=1}^{p} \beta_j^2$ is L-2 norm penalty.

→ Ridge method is useful when ① indep. variables are highly correlated (Multicollinearity) (Under multicollinearity, small change in data → large change in model, making model highly sensitive. Ridge regression stablizes the estimates.) ② When n is large & p is small.

→ Choice of λ: $\lambda \uparrow$, $\beta_j \to 0$. Similar to LASSO, λ should be found using cross validation (find best λ that minimizes cross-validation MSE).

IV. ElastNet Regularization Regression: (Particularly useful when there are multiple features correlated to one another and when there are more X_{ij}'s than observations).

X_{ij}'s

→ ElastNet is linear combination of Lasso and Ridge.

$\min_{\beta_j} \left[\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij} \beta_j)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2 \right) \right]$ α: parameter that balances the penalty terms of LASSO & Ridge. λ: tuning penalty of two methods, $\lambda \uparrow$ → more regularization.

→ Similar to LASSO & Ridge, λ should be found using cross-validation

→ Employer's Requirement of ElastNet: $\min_{\beta_j}$ [objective function] subject to ① Long-only position for all assets and/or ② sum-to-1 constraints (total asset weight = 100%, prevents leveraged portf.)

entire budget is used w/o excess or shortfall. $\sum_{i=1}^{n} w_i = 1$.

! Use this method to find the closest tracking portfolios given a benchmark/explanatory assets.