LASSO Regression

XI(×) – X's are not independent (are correlated)

LASSO Regression

(Least Absolute Shrinkage and Selection Operator)

- Again, **OLS** finds regression coefficients that **minimize** the SSE:

$$SSE = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - etc.)^2$$

- **LASSO** regression finds coefficients that **minimize**:

$$SSE(L) = SSE + \text{shrinkage penalty} = SSE + \lambda\,(|\beta_1| + |\beta_2| + |\beta_3| + etc.)$$

- That is, the penalty λ is applied over the sum of the **absolute values** of the coefficients, **rather** than over the sum of their **squared values**
- The effect is **similar to Ridge** regression:
  - If we set $\lambda = 0$, LASSO minimizes **SSE** → same as **OLS**
  - If we set $\lambda = \infty$, LASSO yields the **null model** $y = \beta_0$
  - Again, the goal is to **select** the λ that **minimizes** the **Test MSE**
- One important and interesting **difference**: mathematically, the **Ridge coefficients** can **never** be shrunk to **0** (except when $\lambda = \infty$), but some **LASSO coefficients** do become exactly **0** eventually as λ increases
  → **LASSO** falls **in between Subset Selection** and **Ridge**.

When/How to use LASSO

- When and how to use LASSO are very **similar to Ridge**
- **LASSO coefficients** have similar properties than Ridge's → **biased, low variance, scale variant**, etc.
- The only difference is that some LASSO **coefficients** become **0** when λ is sufficiently **large**
- As a consequence, LASSO can be thought off as a **hybrid** between **variable selection** and **shrinkage**
- If it is **not** important to **retain all available** variables in the model, LASSO may be a better choice
- Like with Ridge, i is standard practice to:
  - ✓ **Standardize** the predictors in LASSO models
  - ✓ **Compare** LASSO with several **λ's** to other models (e.g., OLS, Ridge) with **cross-validation** measures of the **Test MSE**

R Tips

`Glmnet(){glmnet}` → Package for Ridge and **LASSO** regressions

The LASSO regression is specified identically to the Ridge regression model, except for **`alpha=1`**

`LASSO.fit=glmnet(X,Y,`**`alpha=1, lambda=0`**`)` → **`alpha=1`** fits a **LASSO** regression; **`lambda=0`** fits an OLS regression (i.e., no shrinkage)

`LASSO.fit=glmnet(X,Y,alpha=1,` **`lambda=1000`**`)` → a lot of shrinkage

`LASSO.fit=glmnet(X,Y,alpha=0,` **`lambda=1000000`**`)` → as lambda gets very large (approaches ∞) most coefficients are shrunk thus yielding a **null** model (i.e., just the **intercept**)

```
LASSO.fit=glmnet(X,Y,alpha=1,
                 lambda=c(0,10,100,1000, 1000000))
```
→ run multiple values of `lambda`

`coef(LASSO.fit)` → Lists all ridge coefficients sorted from the largest to lowest `lambda`

KOGOD SCHOOL
of
BUSINESS