## Degree-of-Freedom of Ridge Regression

- Can we say the complexity of the ridge regression model, which returns a $p$-dim coefficient vector $\hat{\boldsymbol{\beta}}^{\text{ridge}}$, is $p$?
- Although $\hat{\boldsymbol{\beta}}^{\text{ridge}}$ is $p$-dim, the ridge regression doesn't seem to use the full strength of the $p$ covariates due to the shrinkage.
- For example, if $\lambda$ is VERY large, the df of the resulting ridge regression model should be close to $0$. If $\lambda$ is $0$, we are back to a linear regression model with $p$ covariates.
- So the df of a ridge regression should be some number between $0$ and $p$, decreasing wrt $\lambda$.

One way to measure the degree of freedom (df) of a method is

$$df = \sum_{i=1}^{n} \mathsf{Cor}(y_i, \hat{y}_i).$$

Suppose a method returns the $n$ fitted value as $\hat{\mathbf{y}} = \mathbf{A}_{n \times n}\mathbf{y}$ where $\mathbf{A}$ is an $n$-by-$n$ matrix not depending on $\mathbf{y}$ (of course, it depends on $\mathbf{x}_i$'s). Then

$$df = \sum_{i=1}^{n} \mathsf{Cor}(y_i, \hat{y}_i) = \sum_{i=1}^{n} A_{ii} = \mathsf{tr}(\mathbf{A}).$$

For example, for a linear regression model with $p$ coefficients, we all agree that the degree of freedom is $p$. If using the formula above we have

$$df = \mathsf{tr}(\mathbf{H}) = p, \quad \hat{\mathbf{y}}_{\mathsf{LS}} = \mathbf{H}\mathbf{y}$$

which also gives us $df = p$.

For ridge regression, we have $\hat{\mathbf{y}}_{\text{ridge}} = \mathbf{S}_\lambda \mathbf{y}$, where

$$\mathbf{S}_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T = \sum_{j=1}^{p} \frac{d_j^2}{d_j^2 + \lambda} \mathbf{u}_i \mathbf{u}_i^T.$$

We can define the effective df of ridge regression to be

$$df(\lambda) = \text{tr}(\mathbf{S}_\lambda) = \sum_{j=1}^{p} \frac{d_j^2}{d_j^2 + \lambda}.$$

When the tuning parameter $\lambda = 0$ (i.e, no regularization), $df(\lambda) = p$; when $\lambda$ goes to $\infty$, $df(\lambda)$ goes to $0$.

Different from other variable selection methods, the df for ridge regression can vary continuously from $0$ to $p$.