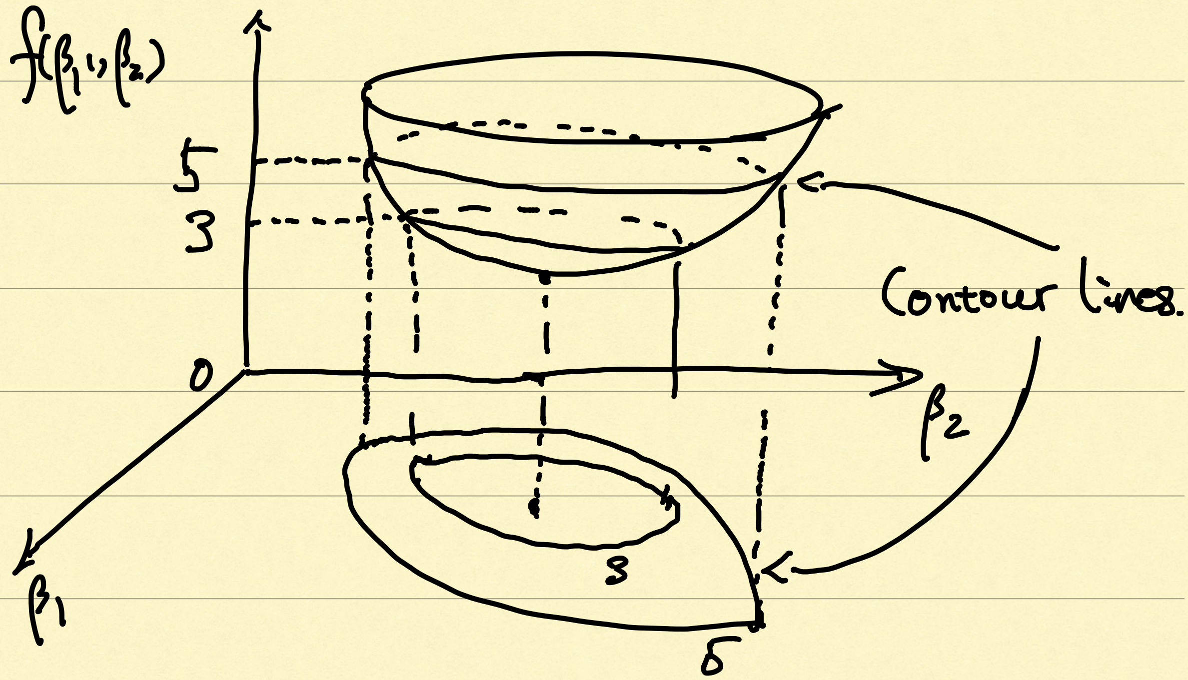

Question: How shall we select the tuning parameter λ?

Model Selection (Each λ index a model).

Basic Method: (Data Splitting)

We split the data $\mathcal{D} = \{X_1, \cdots, X_n\}$ into two subsets $\mathcal{D}_1$ and $\mathcal{D}_2$, such that $\mathcal{D}_1 \cup \mathcal{D}_2 = \mathcal{D}$ with sizes $n_1 + n_2 = n$.

(For example, $n_1 = n_2 = \frac{n}{2}$).

Consider a pool of tuning parameters.

$$\Lambda = \{\lambda_1, \cdots\cdots, \lambda_K\},$$

Let $\hat{\beta}^{\lambda_1}, \ldots, \hat{\beta}^{\lambda_K}$ be the ridge regression estimators on the subset $\mathcal{D}_1$. We define the data splitting score corresponding to λ_k as

$$DS(k) = \frac{1}{n_2} \sum_{i \in \mathcal{D}_2} \left(Y_i - \underset{\sim}{X}_i^T \hat{\beta}^{\lambda_k} \right)^2.$$

We then pick the model with the smallest DS score.

Theory: Conditioning on $\mathcal{D}_1$, it is easy to see that $DS(k)$ is an unbiased estimator of $R(\hat{\beta}^{\lambda_k})$. $\square$

Pro and Cons of data splitting:

Pro: Theoretically and Computationally simple.

Con: "Waste" of training data.

$\Downarrow$

Solution: Cross-validation. (CV)

Def. (J-fold Cross-Validation): We split the data $\mathcal{D}$ into J equal sized parts $\mathcal{D}_1, \ldots, \mathcal{D}_J$. This forms J binary splits.

(DS1) $\mathcal{D}_1$ vs. $\mathcal{D} \backslash \mathcal{D}_1$

(DS_2): $\mathcal{D}_2$ vs. $\mathcal{D} \setminus \mathcal{D}_2$

$\vdots$

(DS_J): $\mathcal{D}_J$ vs. $\mathcal{D} \setminus \mathcal{D}_J$.

For each $\lambda_k \in \Lambda$, we calculate the data splitting scores using $DS_1, \cdots, DS_J$. Denote the results as. $DS_1(k), \cdots, DS_J(k)$ the cross-validation score is $CV(k) := \frac{1}{J} \sum_{j=1}^{J} DS_j(k)$.

We then pick the model with the smallest CV score.

Note: After picking λ, we use the chosen λ to fit on the entire dataset.

Extension from the ridge regression to bridge regression.

Let $\underset{\sim}{x} \in \mathbb{R}^n$, $\|\underset{\sim}{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$.

For $1 \le p < \infty$, $\|\underset{\sim}{x}\|_p$ is a norm.

For $0 < p < 1$, $\|\underset{\sim}{x}\|_p$ is not a norm.