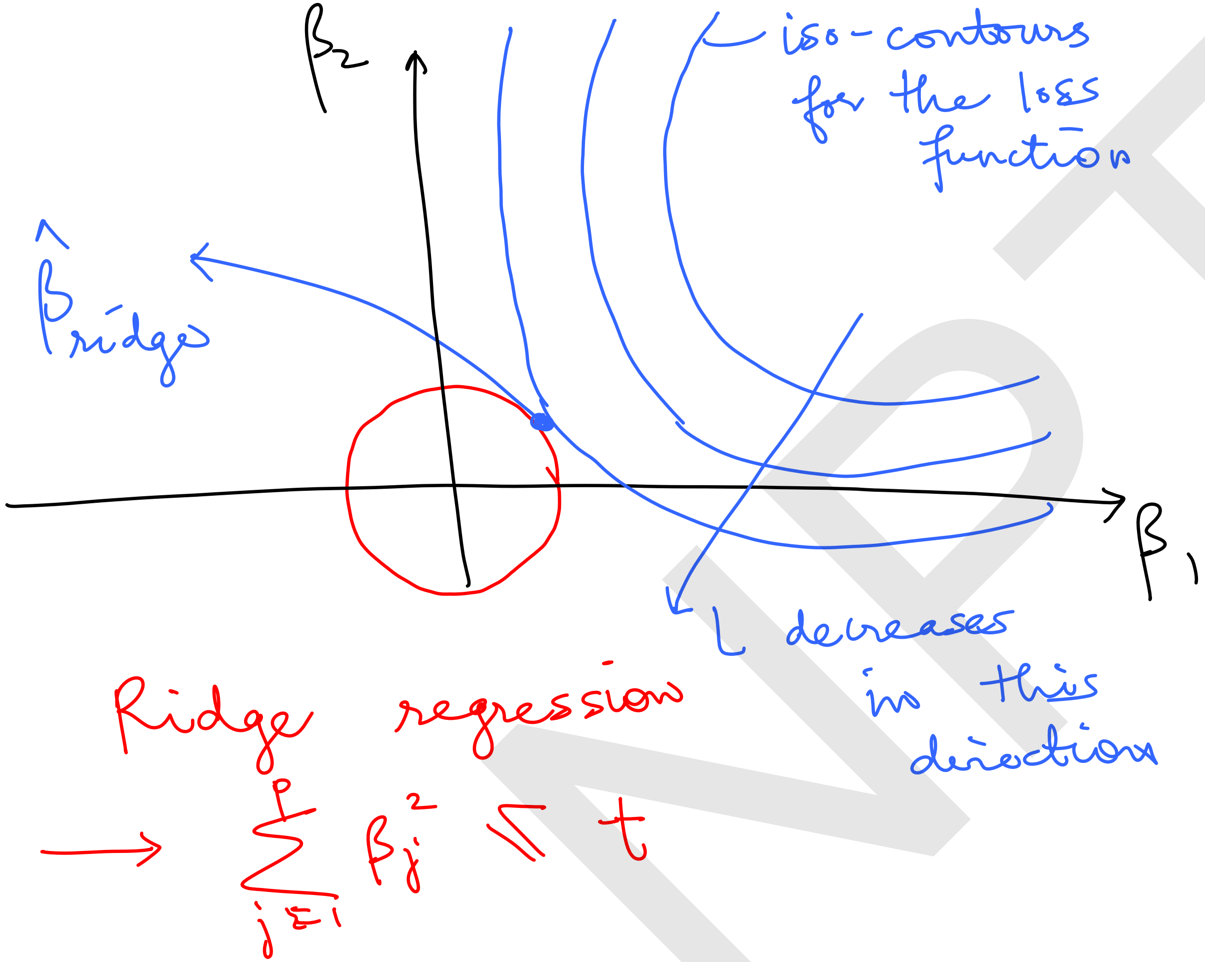
Least absolute shrinkage and selection operator (LASSO)

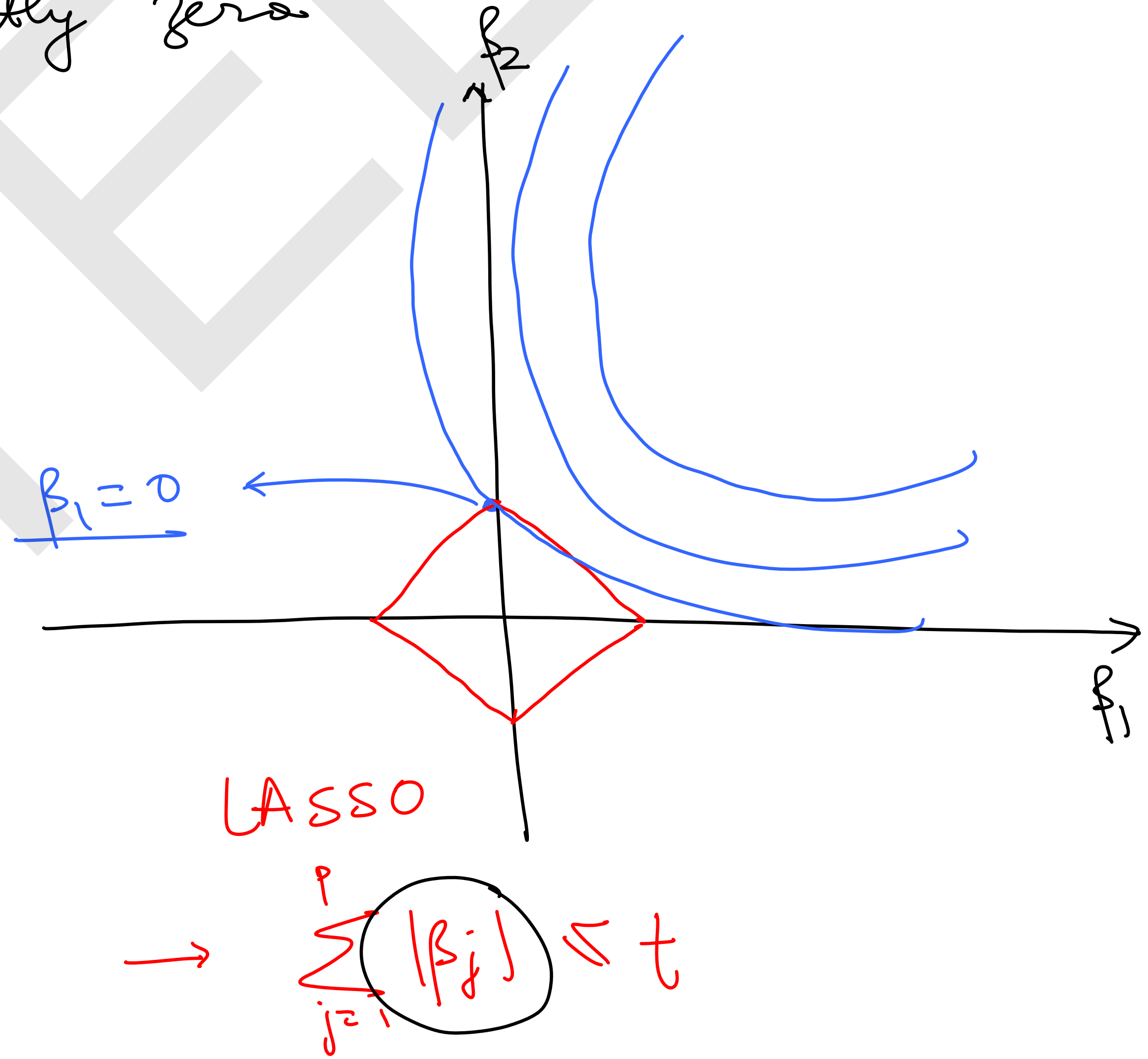
- Least absolute shrinkage refers to the lowest possible values of the parameters in the linear regression model

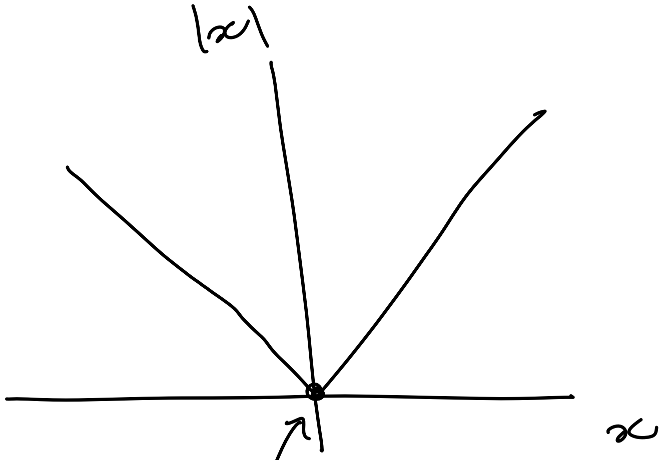
- Selection means that LASSO is able to select the important parameters by making some of the parameters almost zero.

$$L(\beta) = \underbrace{\sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right)\right)^2}_{SSE} + \underbrace{\lambda \sum_{j=1}^{p} |\beta_j|}_{\text{penalty for regul}}$$

As opposed to ridge regression, in LASSO, some parameters in the optimal $\hat{\beta}$ can be exactly zero







function $|x|$ is not differentiable @ $x=0$.

$$L = SSE + \lambda \sum_{j=1}^{p} |\beta_j|$$

function non-analytically differentiable when any parameter β_j is close to zero.

$\Downarrow$

Analytical solution is not possible for LASSO, although we derived one for ridge regression.

→ $\boxed{\text{Numerical methods}}$ can be used to minimize the loss function and thus select optimal $\hat{\beta}_{LASSO}$.

→ λ value should be chosen by $\boxed{\text{cross-validation}}$ to obtain good performance on both train and the test sets.

<u>Elastic nets</u> combines the loss function for ridge regression and LASSO

$$L(\beta) = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \sum_{j=1}^{p}\left(\lambda_2 \beta_j^2 + \lambda_1 |\beta_j|\right)$$

Here, you regularize the loss function using both the sum of squared parameters and sum of absolute values of the parameters.