2.2. Ridge Regression / ℓ^2 regularization.

Setting: given sample points $x^{(1)}, \dots, x^{(n)}$

with $x^{(i)} \in \mathbb{R}^p$.

Form $X = \begin{bmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{bmatrix} \in \mathbb{R}^{p \times n}$

Recall: if $y = f(x; \beta^*) + \epsilon = x^T \beta^* + \epsilon$ is the underlying linear func with noise.

Suppose XX^T is invertible, then

$$\beta^{ls} = (XX^T)^{-1} X y$$

The approximation error is

$$\beta^{ls} - \beta^* = (XX^T)^{-1} X y - \beta^*$$

$$= (XX^T)^{-1} X (X^T\beta^* + \epsilon) - \beta^*$$

$$= \cancel{(XX^T)^{-1} X X^T} \beta^* + (XX^T)^{-1} X\epsilon - \beta^*$$

$$= (XX^T)^{-1} X \epsilon$$

Example: Take $X = \begin{bmatrix} 0.707607 & 0.706607 \\ 0.706607 & 0.707607 \end{bmatrix}$

$$XX^T = \begin{bmatrix} 1 & 0.999999 \\ 0.999999 & 1 \end{bmatrix}$$

and

$$(XX^T)^{-1} X = \begin{bmatrix} 500\,000 & -499\,999 \\ -499\,999 & 500\,000 \end{bmatrix}$$

Observation: the difference $(XX^T)^{-1}X\epsilon$ can be huge for some X. In this case:

(1) β^{ls} is not a reasonable approximation of β^*.

(2) the noise takes too much weight in β^{ls}, i.e. the model fits too much to the noise.

Idea to reduce overfitting:

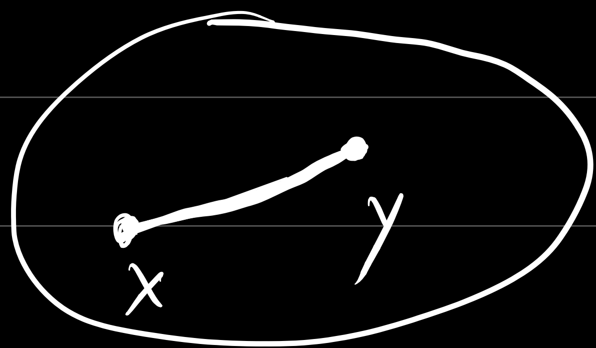
$$\beta^{ridge} = \arg\min_{\beta} \left(\|y - X^T\beta\|^2 + \lambda \|\beta\|^2 \right)$$

where $\lambda \geq 0$ is a constant.

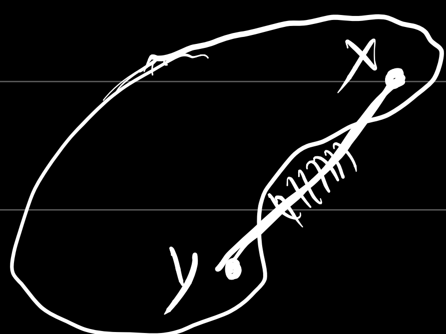
This is called the ridge regression / ℓ^2 regularization / Tikhonov regularization

2.2.1 Math Prep.

Def: A set $C \subset \mathbb{R}^n$ is called <u>convex</u> if $\forall\, x, y \in C$, $\forall\, t \in [0,1]$, we have $tx + (1-t)y \in C$.



C convex



not convex

Lemma: If C_1 and C_2 are convex sets, then $C_1 \cap C_2$ is convex.

Proof: $\forall\, x, y \in C_1 \cap C_2$, $\forall\, t \in [0,1]$

$$C_1 \text{ is convex} \xRightarrow{\text{def of convexity}} tx + (1-t)y \in C_1$$

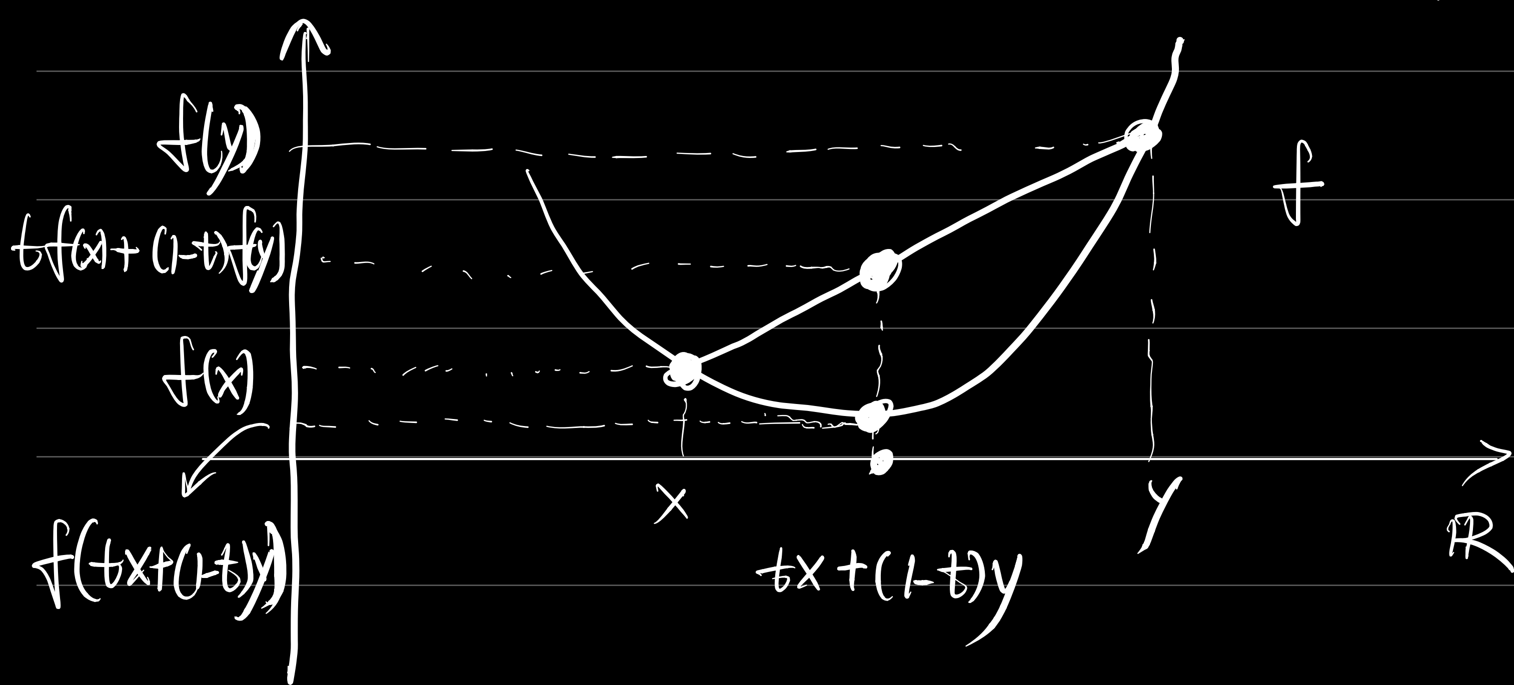
$$C_2 \text{ is convex} \overset{\text{def of convexity}}{\Longrightarrow} tx + (1-t)y \in C_2$$

$$\Rightarrow \quad tx + (1-t)y \in C_1 \cap C_2$$

$$\overset{\text{def of convexity}}{\Longrightarrow} \quad C_1 \cap C_2 \text{ is convex} \quad \blacksquare$$

Def: A func $f: D(f) \to \mathbb{R}$ is called convex if the domain $D(f)$ is convex and $\forall\, x, y \in D(f)$, $\forall\, t \in [0,1]$ we have

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$$



Remark: Geometrically, f is convex means the graph of f lies below the line segment connecting $f(x)$ and $f(y)$.