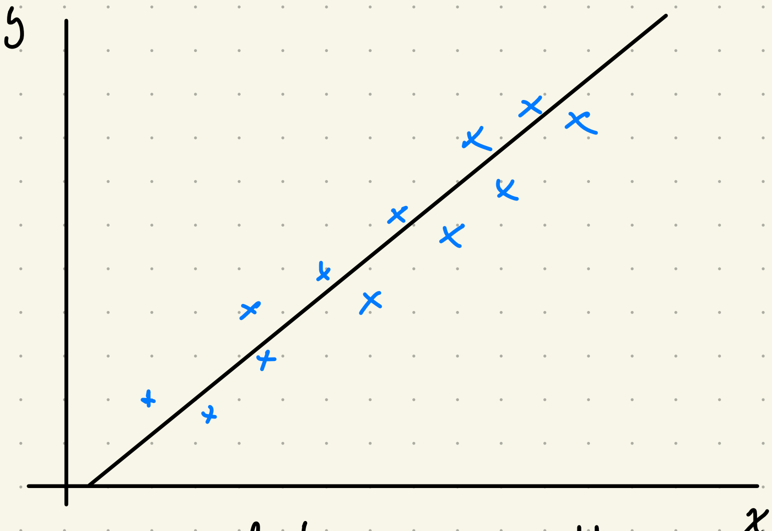
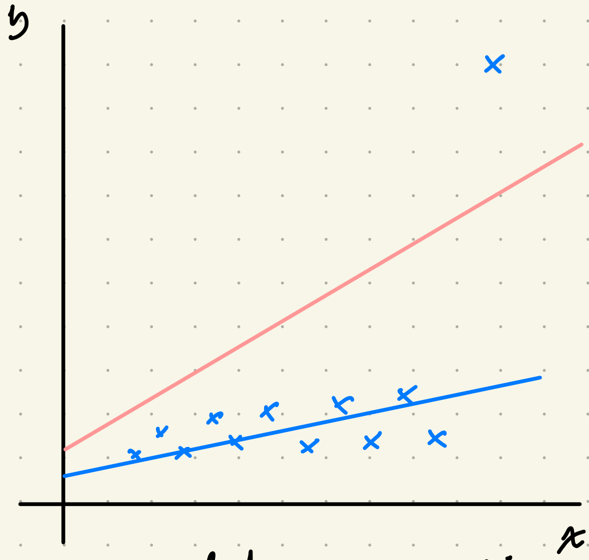
Robustez





$L(y, \hat{y}) = (y - \hat{y})^2 \leftarrow$ función de perdida

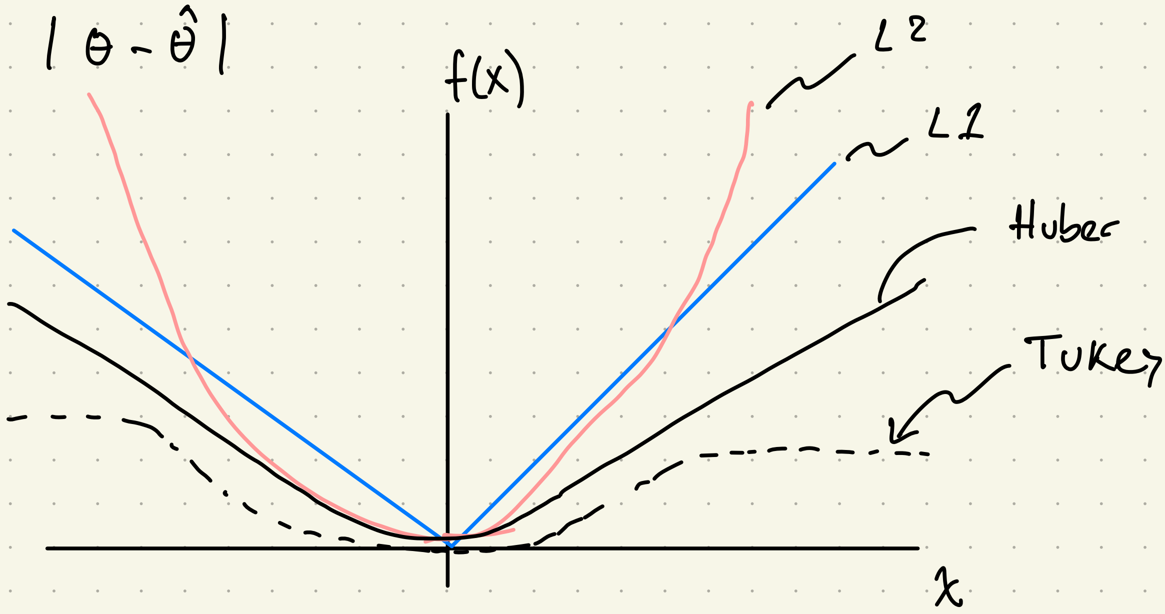
$R(f) = E(L(f(y, f(x))) = E((y - f(x))^2)$

nosotros buscamos minimizar el Riesgo empírico

$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \leftarrow$ func. de perd. cuadrática

norma L2

L1 : $|\theta - \hat{\theta}|$



Regularización

Con regularización buscamos minimizar error de estimación en conjunto con la norma de los estimadores (shrinkeage de parámetros) => disminuir el error de generalización (riesgo empirico).

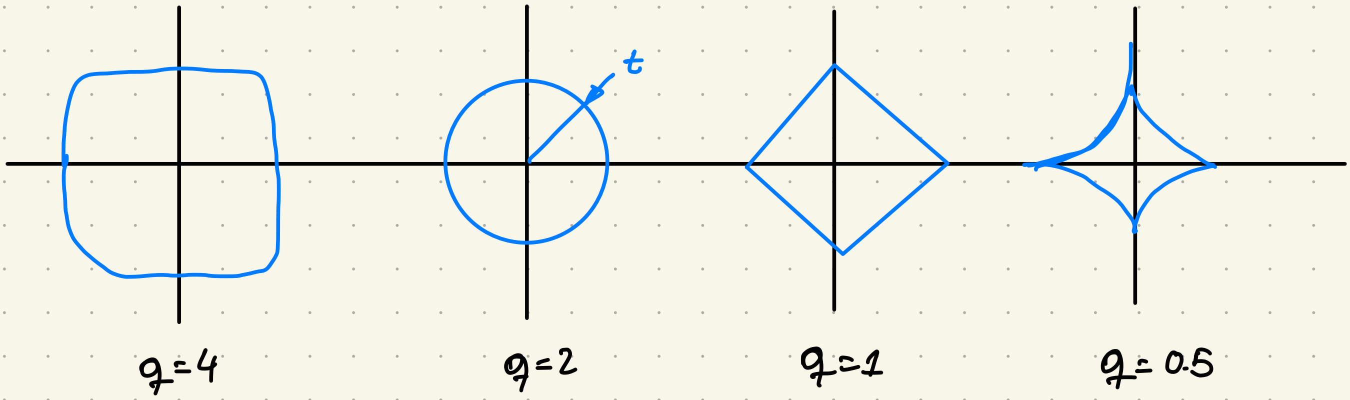
$$\hat{y} = w_0 + w_1 x_1 + \dots + w_m x_m \quad \to \quad \hat{y} = 0.5\, x_1 + 200\, x_2$$

la recta es sensible a x_2

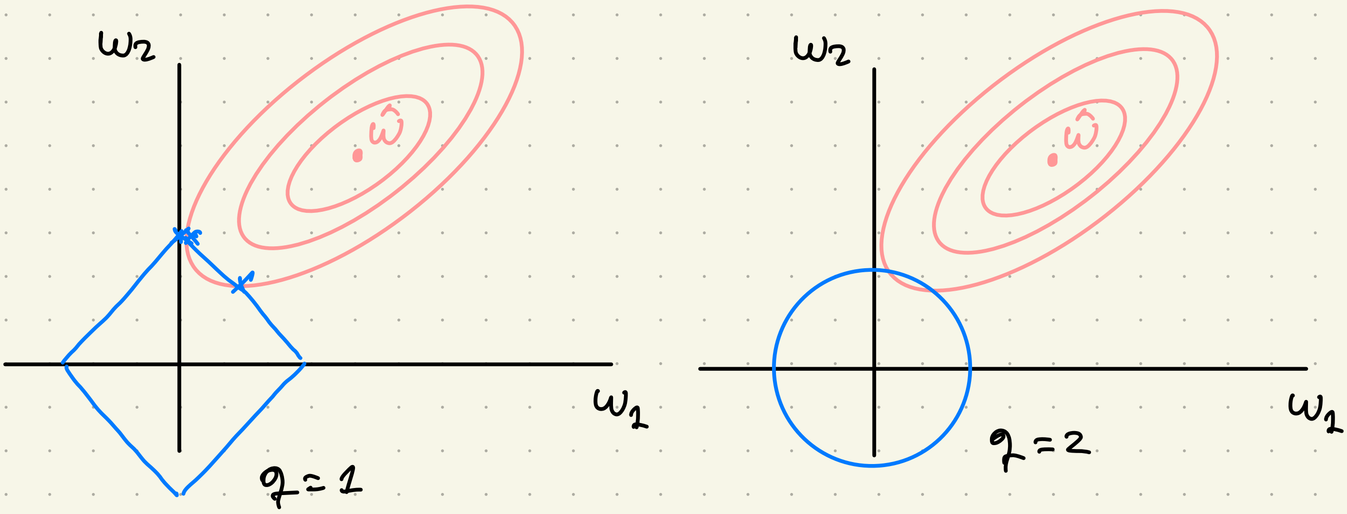
$$\hat{w} = \arg\min_{w} \sum_i \left(y_i - w_0 - \sum_j w_j x_{ij}\right)^2 , \quad \text{con restricción} \quad \underbrace{\sum_j (w_j)^2}_{\|w_j\|^2} \leq t$$

$$\hat{w}^q = \arg\min_{w} \left(\sum_{i=1}^{n} y - \left(\sum_{j=0}^{p} w_j \cdot x_{ij}\right)\right)^2 + \lambda \cdot \sum_{j=0}^{p} w_j^q$$

Gráficamente la restricción determina la región de posibles $\hat{w}_j$. la forma de esta región viene dado por q.



Las curvas de nivel de la función a minimizar son elipses, esto viene del estimador de w. Nosotros buscamos el punto donde la restricción corta la elipse. Si consideramos dim 2 (tamaño de parametros). graficamos $q=1,2$:



si $q=1$, la intersección puede estar en las esquinas $\Rightarrow$ $w_j = 0$ $\Rightarrow$ produce una selección de variables $\Rightarrow$ $q=1 \rightarrow$ reg. L1 o Lasso.

$q=2 \rightarrow$ reg L2 o ridge.