ℓ_2-regularization

$$\text{loss}(y, \hat{y}) = (y - \hat{y})^T (y - \hat{y})$$

$$= (y - X.w)^T (y - Xw) + \lambda w^T . w$$

$$\frac{\partial\, \text{loss}}{\partial w} = -2 . X^T . (y - X.w) + 2 . \lambda . w$$

$$-\cancel{2}X^T . y + \cancel{2} . X^T . X . w + \cancel{2} . \lambda^{I}\, w = 0$$

$$(X^T . X + \lambda I)^{-1} (X^T . X + \lambda . I) . w = X^T . y$$

$$(X^T . X + \lambda I)^{-1}$$

No Regularization

$$w^* = \boxed{(X^T \cdot X)^{-1}} \cdot X^T \cdot y \qquad X = \begin{bmatrix} | & | \end{bmatrix}$$

l_2- regularization with λ parameter.

ridge or tikhonov regularization,

$$w^* = \boxed{(X^T \cdot X + \lambda I)^{-1}} \cdot X^T \cdot y$$

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{D \times D} + \begin{bmatrix} \lambda & & & 0 \\ & \lambda & & \\ & & \lambda & \\ 0 & & & \ddots \; \lambda \end{bmatrix}_{D \times D}$$

$\lambda \uparrow \quad w \downarrow$

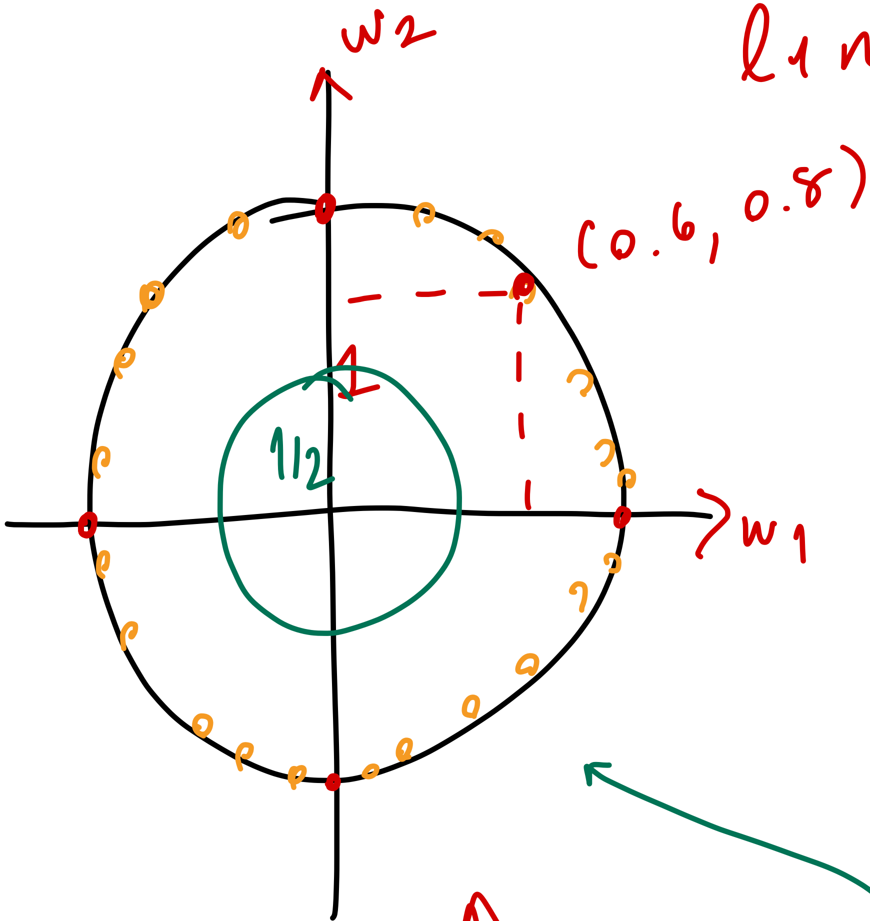
$$\text{loss}(y, \hat{y}) + \underbrace{\lambda \sum_{d=1}^{D} |w_d|}_{\ell_1\text{-regularization.}}$$

$$\ell_p\text{-norm} = \left[\sum_{d=1}^{D} (|w_d|)^p \right]^{1/p}$$

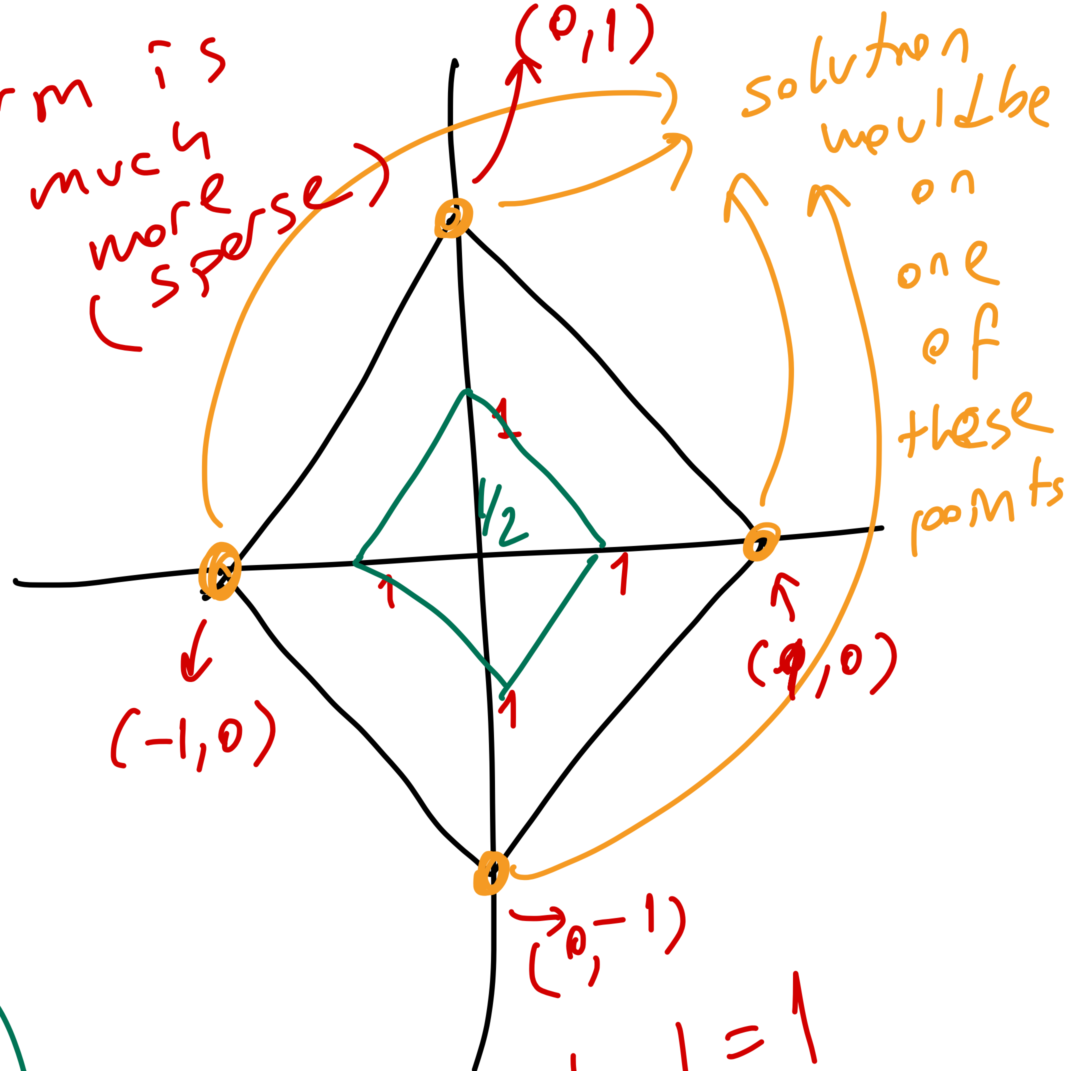
$p \to 4/3$

$$\underline{\underline{p=2}} \Rightarrow \sqrt{w_1^2 + w_2^2 + \dots + w_D^2}$$

$$p=1 \Rightarrow |w_1| + |w_2| + \dots + |w_D|$$



l_1 norm is much more (sparse)



$w_1^2 + w_2^2 = 1$

$w_1^2 + w_2^2 = 1/4$

$|w_1| + |w_2| = 1$

$|w_1| + |w_2| = 1/2$