Supervised Learning - Regularization

Adapting to overfitting

$\ell_p, \ell_{p,q} - norms$

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{n} x_i^p \right)^{\frac{1}{p}}.$$

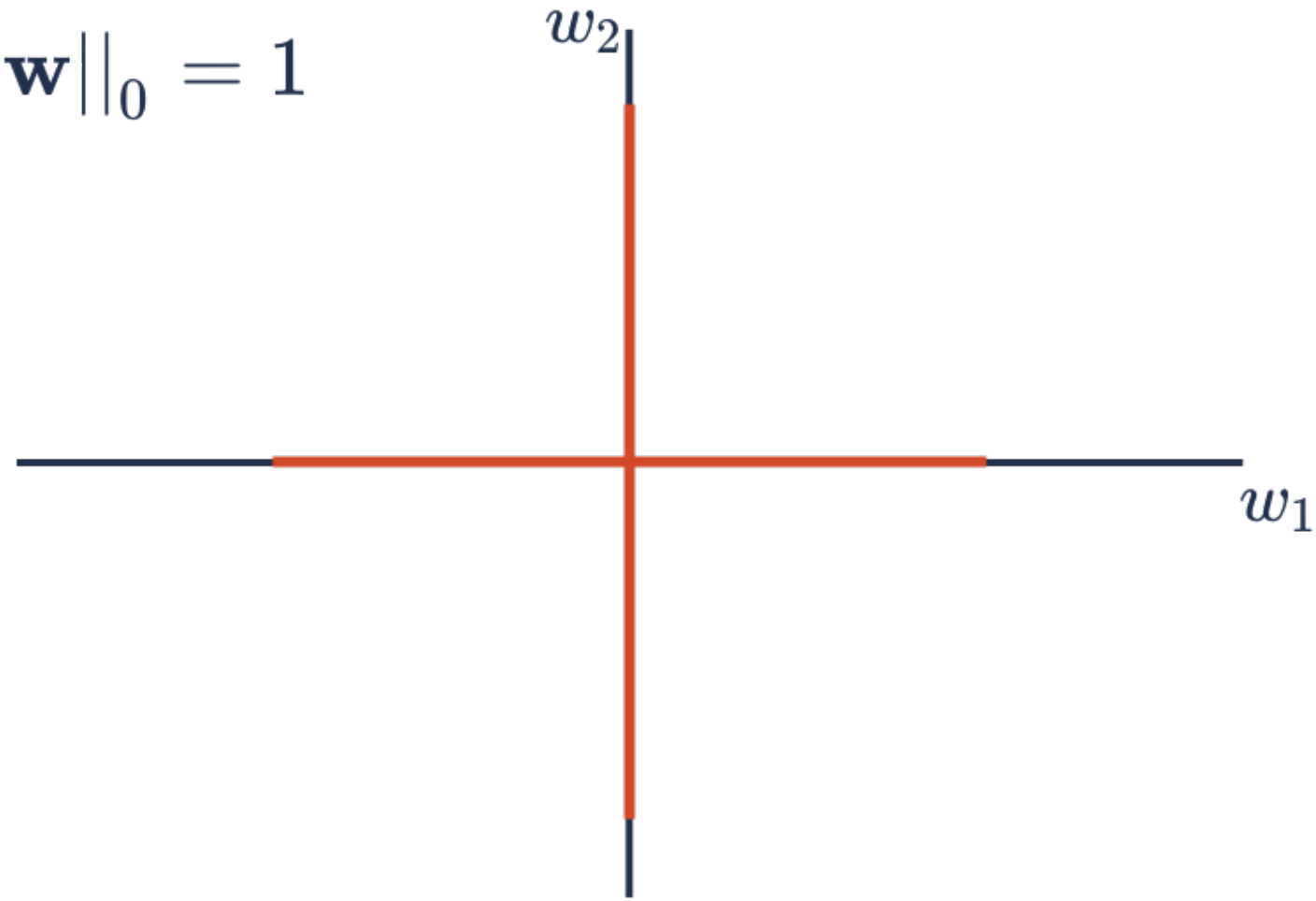
$$||\mathbf{X}||_{p,q} = \left(\sum_{j=1}^{n} ||\mathbf{x}_j||_p^q \right)^{\frac{1}{q}}.$$

$$||\mathbf{X}||_{p,q} = \left(\sum_{j=1}^{n} ||\mathbf{x}_j||_p^q \right)^{\frac{1}{q}} = \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} x_{ij}^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}.$$

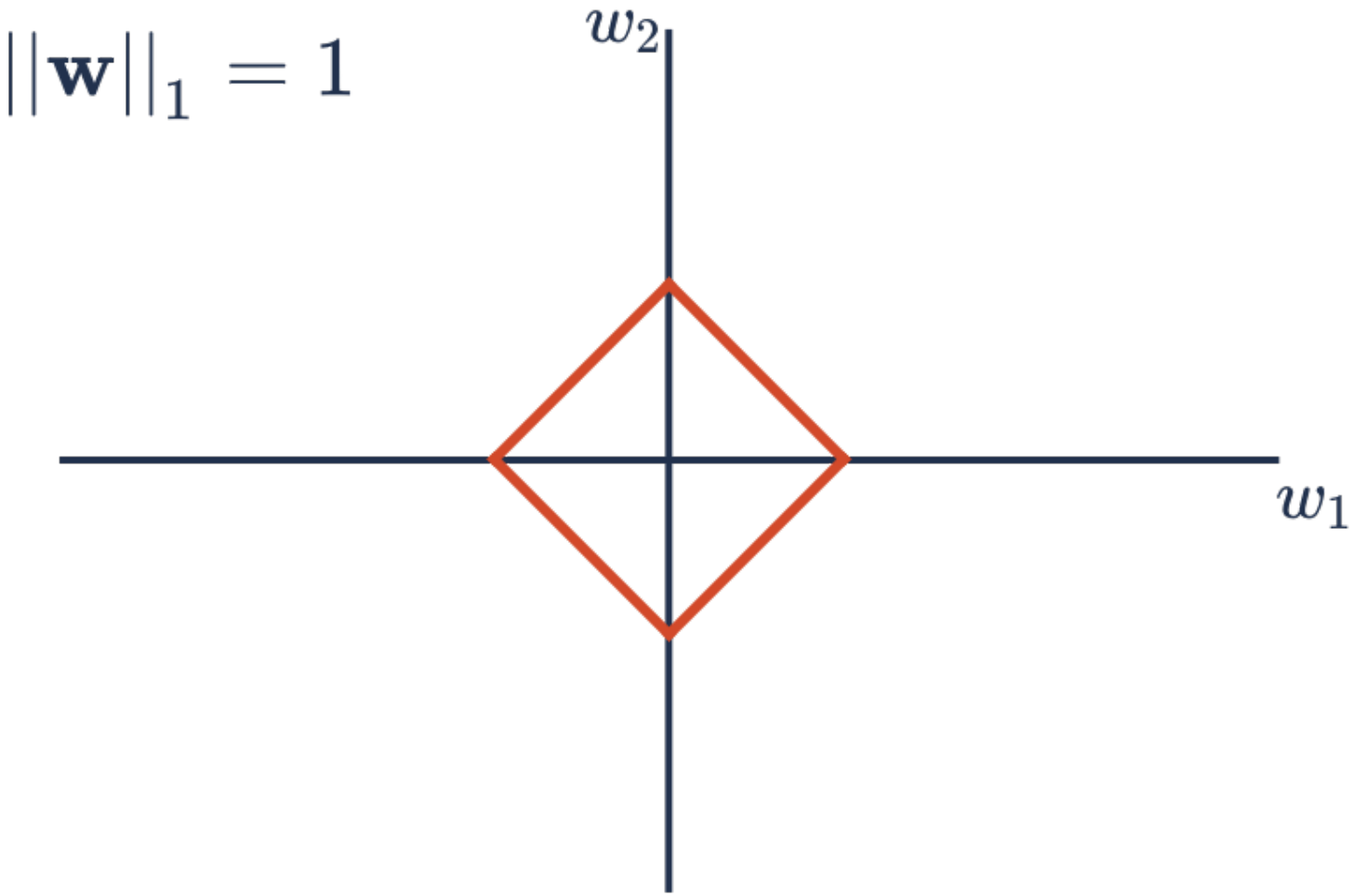
$$||\mathbf{X}||_F = ||\mathbf{X}||_{2,2}.$$

ℓ_0-norm

$$||\mathbf{w}||_0 = 1$$

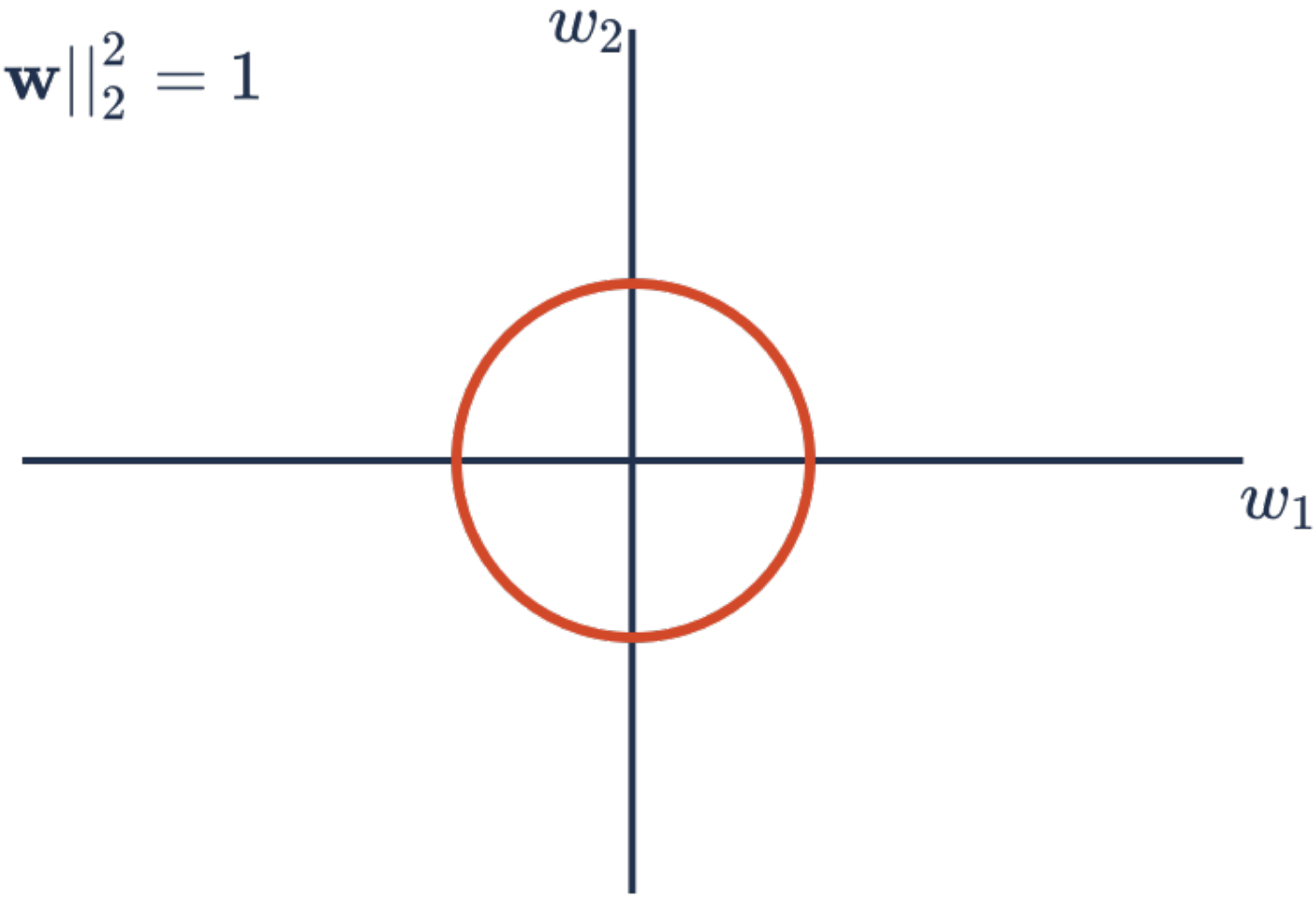


ℓ_1-norm

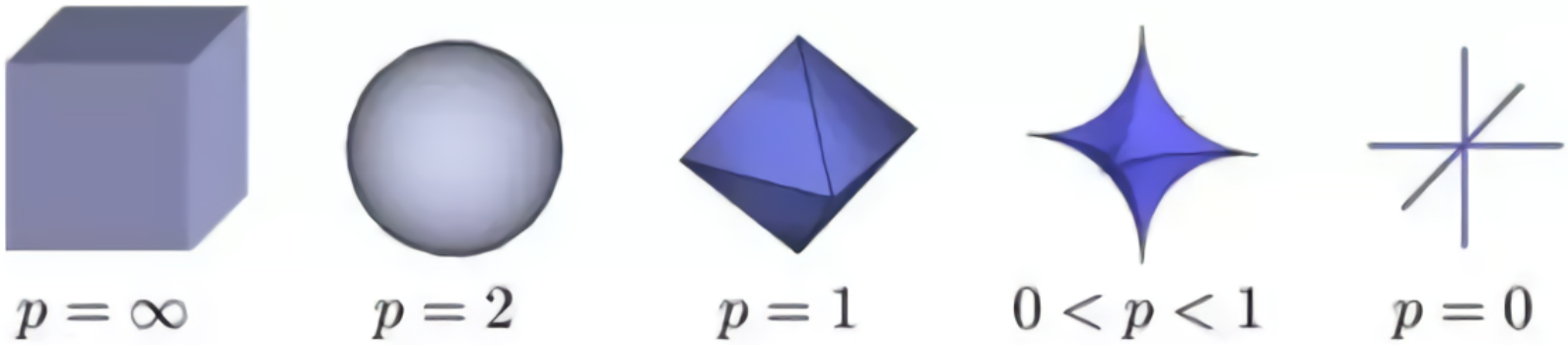


ℓ_2-norm

$$||\mathbf{w}||_2^2 = 1$$



ℓ_p-norms



Linear Regression & Regularizations

Linear Regression

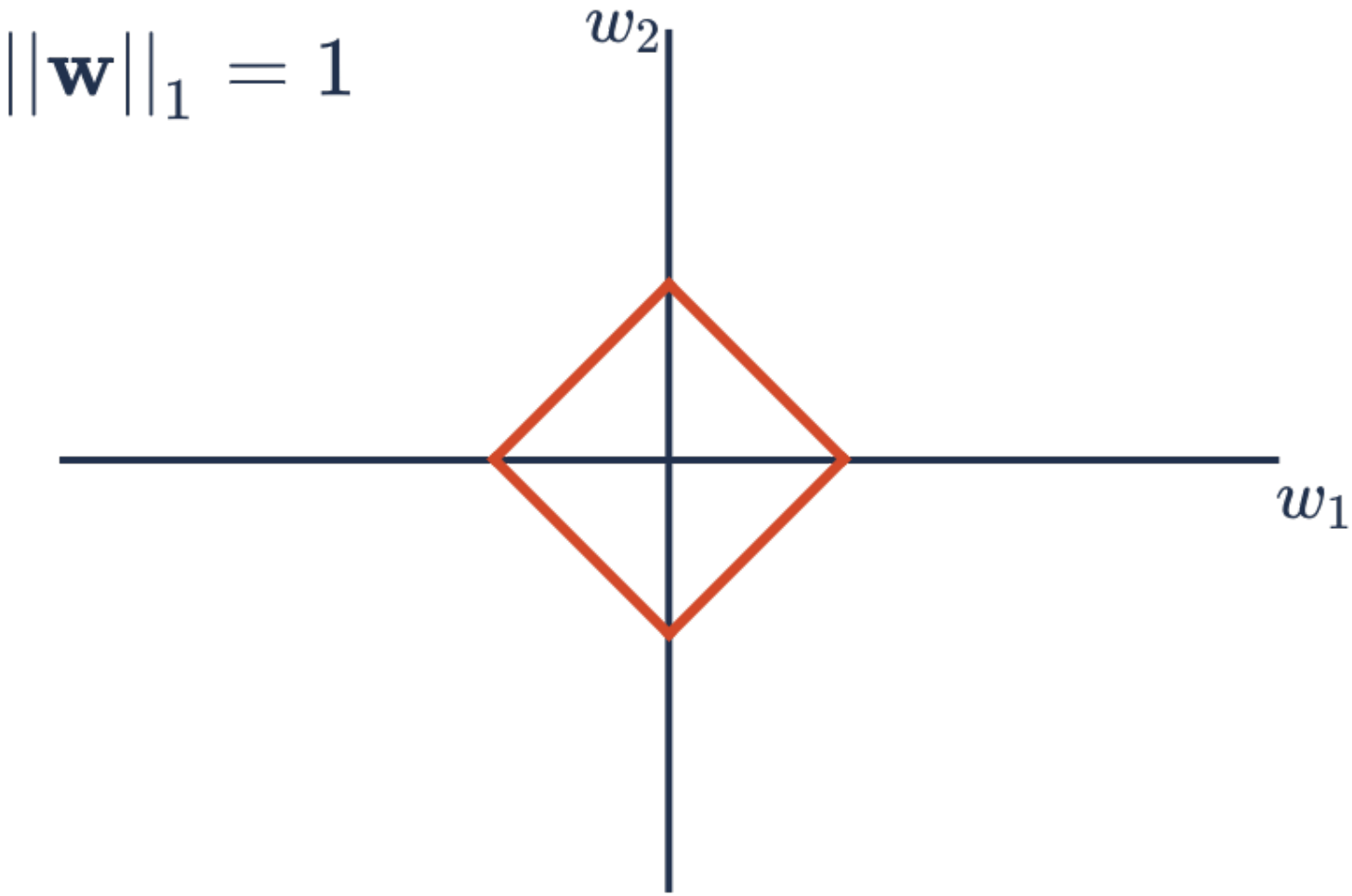
$$\min_{\mathbf{w}} ||\mathbf{y}^T - \mathbf{w}^T\mathbf{X}||_2^2.$$

Lasso

$$\min_{\mathbf{w}} ||\mathbf{y}^T - \mathbf{w}^T\mathbf{X}||_2^2 + \alpha||\mathbf{w}||_1,$$

where α is a constant hyperparameter of our model.

ℓ_1-norm



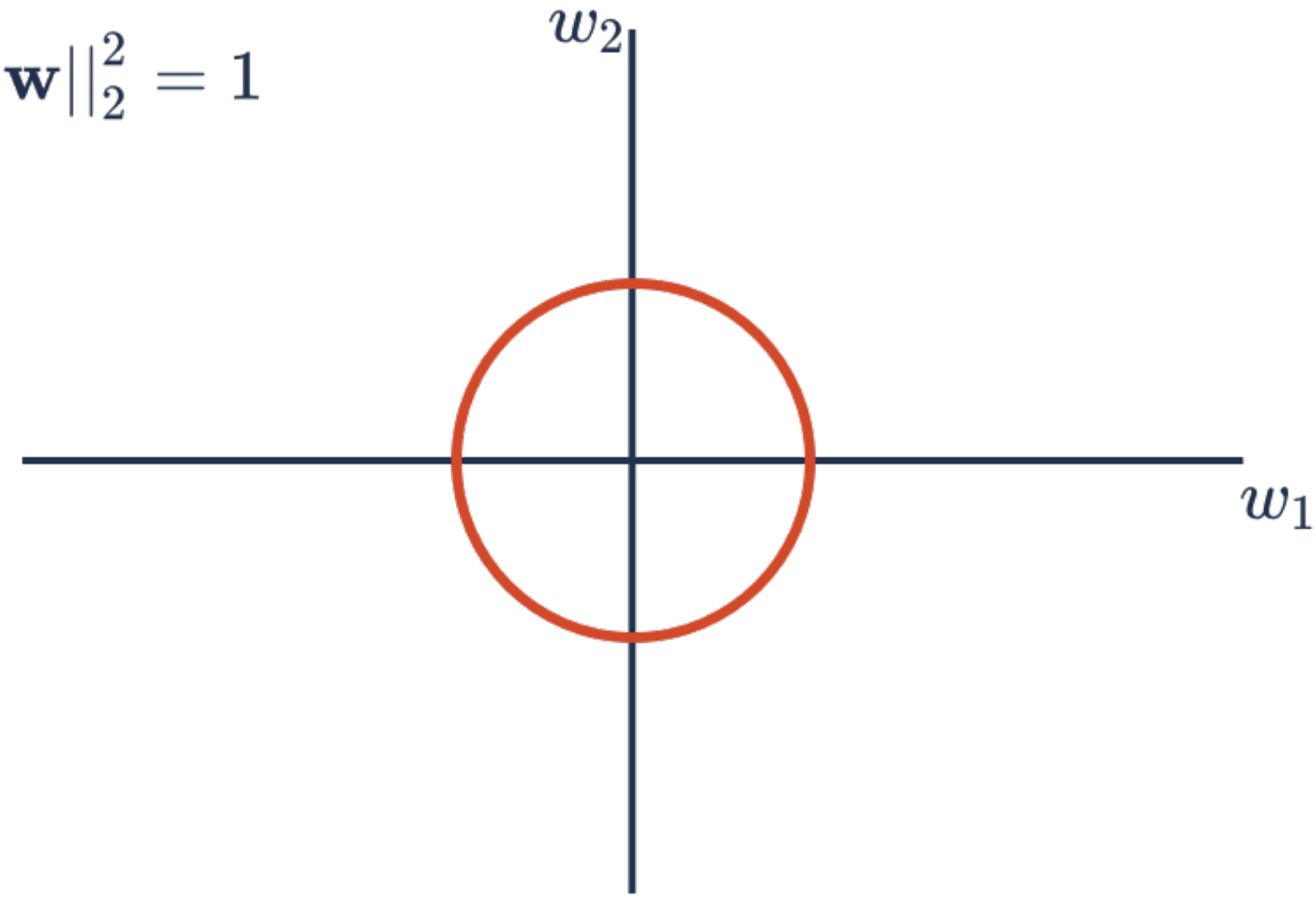
Ridge Regression

$$\min_{\mathbf{w}} ||\mathbf{y}^T - \mathbf{w}^T\mathbf{X}||_2^2 + \alpha||\mathbf{w}||_2^2,$$

where α is a constant hyperparameter of our model.

ℓ_2-norm

$$||\mathbf{w}||_2^2 = 1$$



Elastic Net

$$\min_{\mathbf{w}} ||\mathbf{y}^T - \mathbf{w}^T\mathbf{X}||_2^2 + \lambda_1||\mathbf{w}||_1 + \lambda_2||\mathbf{w}||_2^2,$$

where λ_1, λ_2 are the coefficients for the ℓ_1 and ℓ_2-norms respectively.

Elastic Net

$$\min_{\mathbf{w}} ||\mathbf{y}^T - \mathbf{w}^T\mathbf{X}||_2^2 + \alpha\rho||\mathbf{w}||_1 + \frac{\alpha(1-\rho)}{2}||\mathbf{w}||_2^2.$$

where α is the normalization coefficient and ρ is a balancing ratio between the ℓ_1 and ℓ_2-norms.

Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.