Regularization (Shrinkage)

As the name Suggest this section is about **Regularizing** or **Constraints** the coefficients estimates $\hat{\theta}$ also noted as $\hat{\beta}$, These are the two methods discussed here :

- Ridge Regression L_1
- The Lasso L_2

Before diving into these methods, taking a look at the **Norms** will help understanding and intuition since they are derived from.

Norms

When thinking of geometric vectors intuitively the direction and length of the vector are first that comes to mind, Simply **Norm** is a function that assigns each vector x it's **length** $\|x\|$ or **magnitude**

- $\|\lambda x\| = |\lambda| \|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|x\| \geq 0$ and if $\|x\| = 0 \iff x = 0$

The L_p Norm

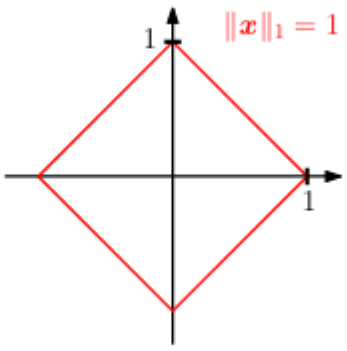
Also written as $\|x\|_p$, is defined as:

$$\|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

with : $p > 0$ and x_i the **components** of x

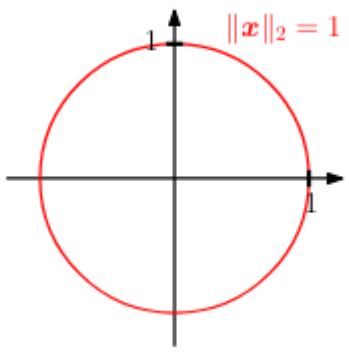
The L_1 Norm (Manhattan Norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$



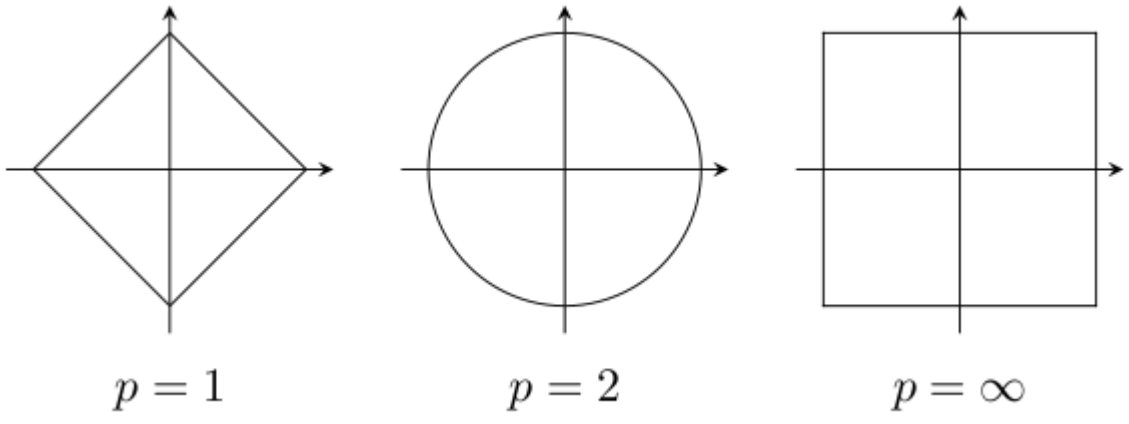
The L_2 Norm (Euclidean Norm)

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$



The L_∞ Norm

$$||x||_\infty = \max_i(|x_i|)$$



- Which results in a **square**