Regularization and the Two Norms

Demo for the Beamer Template

Istiyak H. Siddiquee

Faculty of Computer Science,
Otto-von-Guericke University, Magdeburg.

1 Definition

What is Regularization

Regularization

"Penalizing the weights and biases of the error function to keep them from reaching large values is called Regularization [1]". Mathematically,

$$\underset{x}{\text{minimize}} \qquad \|Ax - b\|_2^2 + \lambda \|x\|_n \tag{1}$$

The last term of equation 1 is the regularization term here.

2 The Two Norms

Two Norms of Interest

There are two most widely used norms that we are interested in:

- The ℓ_1-norm
- The ℓ_2-norm

The ℓ_1-Norm

Replacing n of equation 1 with 1, i.e. using the absolute values will result in ℓ_1-norm. So, the following would be ℓ_1 regularized cost functions:

$$\underset{x}{\text{minimize}} \qquad \|Ax - b\|_2^2 + \lambda \|x\|_1 \tag{2}$$

The $_2$-Norm

Replacing n of equation 1 with 2, i.e. using the squared values of the weights, will result in ℓ_2-norm. So, the following would be ℓ_2 regularized cost functions:

$$\underset{x}{\text{minimize}} \qquad \|Ax - b\|_2^2 + \lambda \|x\|_2 \tag{3}$$

2 The Two Norms

Comparison Between the Two Norms

Comparison

	ℓ_1-Norm	ℓ_2-Norm
Sparsity (See Fig 1)	Yes	No
Robustness	Yes	No

Table 1: Differences between the two norms

Sparsity of ℓ_1-Norm

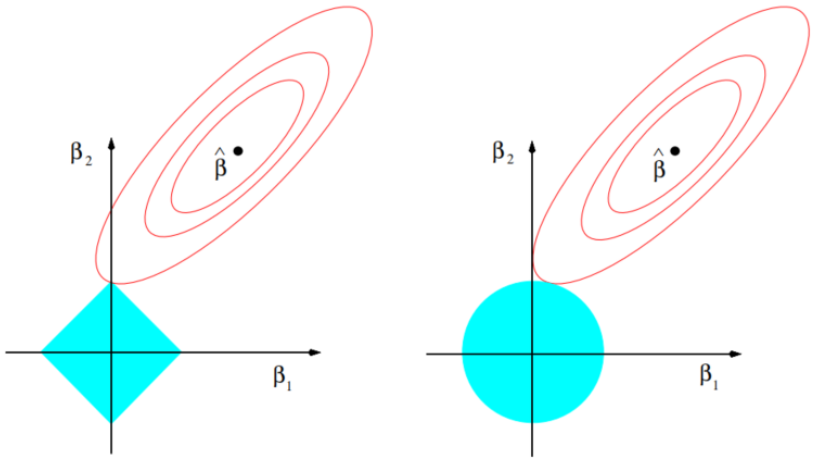

Figure 1: Contours of error and constraint functions for Lasso (Left) and Ridge Regression (Right) [2].

3 References

References I

[1] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0387310738.

[2] J. Friedman, T. Hastie, R. Tibshirani, et al. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York, 2001.

Thank you!