From https://stats.stackexchange.com/questions/191603/ridge-regression-formulation-as-constrained-versus-penalized-how-are-they-equiv

The classic [Ridge Regression](https://en.wikipedia.org/wiki/Tikhonov_regularization) ([Tikhonov Regularization](https://en.wikipedia.org/wiki/Tikhonov_regularization)) is given by:

$argmin\_{x}\frac{1}{2}∥x-y∥\_{2}^{2}$+λ$∥x∥\_{2}^{2}$

The claim above is that the following problem is equivalent:

$$argmin\_{x}\frac{1}{2}∥x-y∥\_{2}^{2}$$

subject to $∥x∥\_{2}^{2}\leq t$

Let's define  $\hat{x}$ as the optimal solution of the first problem and $\tilde{x}$ as the optimal solution of the second problem.

The claim of equivalence means that ∀ *t*, ∃ λ≥0: $\hat{x}$ =$\tilde{x}$ .

Namely you can always have a pair of *t* and λ≥0 such the solution of the problem is the same.

How could we find a pair?

Well, by solving the problems and looking at the properties of the solution.

Both problems are Convex and smooth so it should make things simpler.

The solution for the first problem is given at the point the gradient vanishes which means:

$$\hat{x}-y+2λ \hat{x}=0$$

The [KKT Conditions](https://en.wikipedia.org/wiki/Karush%E2%80%93Kuhn%E2%80%93Tucker_conditions) of the second problem states:

$$\tilde{x}-y+2μ \tilde{x}=0$$

and

$$μ\left(∥\tilde{x} ∥\_{2}^{2}-t\right)=0$$

The last equation suggests that either μ=0 or $∥\tilde{x} ∥\_{2}^{2}=t$.

Pay attention that the 2 base equations are equivalent.
Namely if $\hat{x}$ =$\tilde{x}$  and μ=λ  then both equations hold.

On one hand, if $∥\tilde{x} ∥\_{2}^{2} \leq t$ , then one must set μ=0 which means that for *t* large enough, the two two base equations can be equivalent only if one sets λ=0.

On the other hand, if $∥\tilde{x} ∥\_{2}^{2 }\geq t $, one should find μ such that

$$y^{t}\left(I+2μ I\right)^{-1}\left(I+2μ I\right)^{-1}y=t$$

This is basically when $∥\tilde{x} ∥\_{2}^{2} =t$.

Once you find that μ , the solutions will concide.

Regarding the $L^{1}$ case, well, it works with the same idea.

The only difference is we don't have closed for solution hence deriving the connection is trickier.

Have a look at my answer at [StackExchange Cross Validated Q291962](https://stats.stackexchange.com/questions/291962) and [StackExchange Signal Processing Q21730 - Significance of λλ in Basis Pursuit](https://dsp.stackexchange.com/a/48283/128).

[Share](https://stats.stackexchange.com/a/349056)

Cite

[Improve this answer](https://stats.stackexchange.com/posts/349056/edit)

Follow

answered May 30 '18 at 20:01

[Royi](https://stats.stackexchange.com/users/6244/royi)