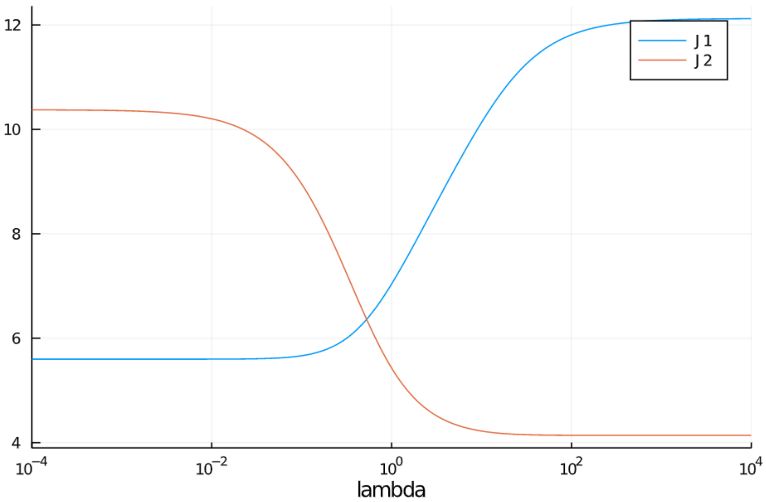
Multi-objective Least Squares

Sometimes, we would like to know about several y-variables' changes when the x-variable changed which is the reason why the multi-objective least squares function is used. For example, suppose we have several objectives $\|A_1x - b_1\|^2, \ldots, \|A_kx - b_k\|^2$. where A_i is an $m_i \times n$ matrix and b_i is an m_i-vector.we seek one x that makes all k objectives small. Usually, no single value of x which minimizes all objectives simultaneously.

Multi-objective least squares via weighted sum:

A standard method for finding a value of x that gives a compromise in making all the objectives small is to choose x to minimize a weighted sum objective:$\lambda_1\|A_1x - b_1\|^2 + , \ldots , + \lambda_k\|A_2x - b_2\|^2$When k equals to 2 and λ_1 equals to 1, we can plot the Two-objective least squares as below:



$$J_1 + \lambda J_2 = \|A_1x - b_1\|^2 + \lambda\|A_2x - b_2\|^2$$

From the left figure, we can see through the lambda increases, the value of J1 increases and the value of J2 decreases. Hence, there is a trade-off between the J1 and J2. In other words, when lambda is small, we care less about J2, and the objective value is closed to least squares of J1. When the lambda increases, we care more about minimizing the value of J2, so the value of J1 increases.

Weighted least squares is also equivalent to a standard least squares problem:

$$\text{minimize} \quad \left\| \begin{bmatrix} \sqrt{\lambda_1}A_1 \\ \sqrt{\lambda_2}A_2 \\ \vdots \\ \sqrt{\lambda_k}A_k \end{bmatrix} x - \begin{bmatrix} \sqrt{\lambda_1}b_1 \\ \sqrt{\lambda_2}b_2 \\ \vdots \\ \sqrt{\lambda_k}b_k \end{bmatrix} \right\|^2$$

And the solution is below:

$$\begin{aligned} \hat{x} &= (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b} \\ &= (\lambda_1 A_1^TA_1 + \cdots + \lambda_k A_k^TA_k)^{-1}(\lambda_1 A_1^Tb_1 + \cdots + \lambda_k A_k^Tb_k). \end{aligned}$$

Regularization

Sometimes we estimate a set of n values, the entries of n-vector x. The set of m measurements, the entries of an m-vector y. The set and measurements are related by $y = Ax + v$, where v is regarded as measurement noise which is unknown but the value is quite small. The problem is to make a sensible guess to x by given A and y.

If we guess that x has the value of $\hat{x}$. Then, it means we are guessing $v = y - A\hat{x}$. Usually, we would have other information about x. In some cases, there is a strong multicollinearity exists in the set of x. We would make a very very big mistake by utilizing the least squares, because the coefficient estimates may change erratically(overfit in machine learning and the curve probably would not smooth) in response to may be only a little bit change in the data. It happens occasionally in models with large numbers of parameters (may happened in economy research).

In this case, we can change the formula in to $\| Ax - y \|^2 + \lambda \| x \|^2$, which is also called Tikhonov Regularization. The stacked matrix in this case is $\begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix} = \tilde{A}$, $\tilde{A}$ is a matrix always has independent columns and the Gram matrix associated with $\tilde{A}$ is as below:

$$\tilde{A}^T\tilde{A} = A^TA + \lambda I$$

The matrix above is always invertible when $\lambda > 0$. The Tikhonov regularized approximate solution is as below:

$$\hat{x} = (A^TA + \lambda I)^{-1}A^Ty$$

This problem can also be written as a Lagrangian: $(y - A\hat{x})^T(y - A\hat{x}) + \lambda(\hat{x}^T\hat{x} - c)$ *and* $\hat{x}$ *subject to* $\hat{x}^T\hat{x} = c$.

We can just consider λ as the Lagrange multiplier in the formula. If λ equals to 0, it would just an ordinary least square problem.