Lecture 7 - *Constrained Linear Regression*

OBJECTIVE: In this lecture, we will learn that it is possible to come up with different cost functions that are more agressive than ridge in selecting the right input features. We will introduce the **lasso** algorithm and quadratic programming.

Textbook: Pages 64–65.

So far, we have dealt with quadratic (L_2) cost functions:

$$C(\theta) = (Y - X\theta)^T(Y - X\theta) + \delta^2\|\theta\|_2^2$$

where $\|\theta\|_2 \triangleq \left(\sum_i \theta_i^2\right)^{1/2}$. So we are solving

$$\min_{\theta\,:\;\theta^T\theta \leq t} \left\{(Y - X\theta)^T(Y - X\theta)\right\}$$

We could also try other constraint norms such as the L_1:

$$C(\theta) = (Y - X\theta)^T(Y - X\theta) + \delta^2\|\theta\|_1$$

where $\|\theta\|_1 = \sum_i |\theta_i|$, thus yielding the optimisation problem:

$$\min_{\theta\,:\;\sum_i |\theta_i| \leq t} \left\{(Y - X\theta)^T(Y - X\theta)\right\}$$

What should dictate our choice of norm? Let us look at a 2D case.

⋆

⋆ Effect on θ:

While we can solve the quadratic constraint problem analytically (ridge regression), we need to carry out **quadratic programming** so solve the L_1 (LASSO) constraint problem.

Lasso as Quadratic Programming

The quadratic programming problem is cast in the following general form:

$$\min_{\theta:\ A\theta \leq t} \left\{ \frac{1}{2}\theta^T H \theta + f^T \theta \right\}$$

In Matlab, one can find the θ that solve the quadratic programming problem by typing `theta = quadprog(H,f,A,t)`. Let us now look at how to convert the lasso problem to generic quadratic programming.

$\star$