Ridge: Soft & Hard Bounds

Samprit Chakraborty

Roll No.- BS2004

Instructor: Dr. Arnab Chakraborty

Indian Statistical Institute, Kolkata

1 Shrinkage Technique

In least square method we minimize

$$\|\vec{y} - x\vec{\beta}\|^2.$$

In case of ridge regression we can interpret it in this way: we are minimizing $|\vec{y} - x\vec{\beta}\|^2 + \kappa \sum {\beta_j}^2$ over β. here tunning parameter is κ. If we choose a value of κ that corresponds to choose a value of τ^2 in the bayesian set up that corresponds to choose a value of λ in the ad hoc set up. All these things are equivalent.

By adding $\kappa \sum {\beta_j}^2$ with $|y - x\beta\|^2$ we are trying to make $\vec{\beta}$ towards 0.
The samething happens in bayesian set up also, where we put the condition that the betas are more or less centered around 0 an mean is $\overrightarrow{0}$ and the variance τ^2 controlled how far from zero we are allowing to g0, here also it is the same idea. All these techniques are called 'Shrinkage techniques' we are trying to shrink β' s down towards 0 , which. Keeps the variance of the estimates of β' s low. In the case of merely least-square there was no bound on that so it could be very large.

If there is no multicollinearity it will e not be very large but it has the potential of becoming very large. In this case we are always keeping a check on that, it is called shrinkage technique.

2 Interpretation of Hard Bound

Another formulation of the above technique is: minimize $|\vec{y} - x\vec{\beta}\|^2$ over $\vec{\beta}$ subject to condition $\sum {\beta_j}^2 \leq \delta$.

So in the condition distance of β' s from the origin is not too large, there is some hard bound delta, we are not allowing $\sum {\beta_j}^2$ to exceed that hard bound. Here tunning parameter is δ.
Now controlling δ, is same as controlling κ, is same as controlling τ^2, is same as controlling λ.