Video 134 - LASSO

Arnab Chakraborty
Scribe-Arghya sarkar (Roll-bs2021)

October 2022

1 Introduction

In previous videos we mentioned how Ridge regression method is preferred than set selection to handle multicollinearity. It was because set selection throws some variables away so we lose some information. Also it is often difficult to determine which variables we should throw. But sometimes in many cases we collect data about large amount of variables and we want to throw some variable away whose contribution is very little. Now we don't know which variables are insignificant and if we try all subset selection the computation will be too long (since if p is large 2^p will be even larger).

In this situation we use LASSO. This is a variation of Ridge regression which is by far most popular way to deal with multicollinearity in modern application of statistics.

2 Formulation

Formulation of LASSO is similar to hard-bound formulation of Ridge regression. In that case we gave bound on $\vec{\beta}$ as $||\vec{\beta}||_2 \leq \delta$. For LASSO we give bound $||\vec{\beta}||_1 \leq \delta$. So then we need to find suitable solution $\hat{\vec{\beta}}$ of the following optimization problem

$$\min_{\vec{\beta}} ||\vec{y} - X\vec{\beta}||_2^2 \quad \text{subject to } \sum |\beta_j| \leq \delta$$

Now this is not solvable analytically. However efficient computational solution can be achieved for this problem. That's how we can solve it using various software.

3 Advantage of using LASSO

The solution of $\hat{\vec{\beta}}$ that minimize the above optimization problem has one additional property that many of $\hat{\beta}_i$'s will be close to zero. In general, LASSO gives sparser fit than Ridge regression. Since a lot of coefficients of these variables turns out to be zero or very close to zero we can throw away those variables and describe the model with smaller number of variables without any significant loss. This combines the effect of both subset selection and Ridge regression.

Also LASSO is efficient than both Ridge regression and subset selection. Since it gives sparser coefficients than Ridge regression which makes models easier to interpret. On the other hand, it does not need to check all 2^p many possible subset of p variables like subset selection.

4 Application

Note using LASSO a lot of the coefficients becomes very small allowing us to discard their corresponding variables which enables us to describe a model with fewer variables which is always easier to interpret. Specially in modern applications of statistics it is quite common to have thousands of regressor variables. In those cases using LASSO is very useful. That is why it is very popular in modern statistics.