7&8. What is Ridge regression and Lasso regression?

Ridge and Lasso are two shrinkage methods. They shrink the regression coefficients by imposing a penalty on their size. There are two motivations to use Ridge and Lasso regressions:

**Motivation 1: too many predictors**

* It is not unusual to see the number of input variables greatly exceed the number of observations, e.g. micro-array data analysis, environmental pollution studies.
* With many predictors, fitting the full model without penalization will result in large prediction intervals, and LS regression estimator may not uniquely exist.

**Motivation 2: correlated variables X in a linear regression model**

* Because the LS estimates depend upon , we would have problems in computing  if X′X were singular or nearly singular.
* In those cases, small changes to the elements of X lead to large changes in . In other words, due to the correlated variables, the coefficients are poorly determined and exhibit high variance.

The Ridge and Lasso coefficients minimize a penalized residual sum of squares:

Here is a complexity parameter that controls the amount of shrinkage: the larger the value of the greater the amount of shrinkage. The coefficients are shrunk toward zero.

The difference between Ridge and Lasso regression is that when the estimated coefficients have sharp boundary, e.g. the coefficients are piece-wise, Lasso can capture this sharp boundary better than Ridge.