Jan 19th

Meeting record

In today’s meeting, we discuss the shrinkage part (cri 1). From Copas Regression, Prediction and Shrinkage, shrinkage is defined by ‘the amount by which validation fit falls short of retrospective fit‘ and is greatly affected by empirical model selection such as stepwise regression or optimal subset methods. The shrinkage factor of usual multiple regression model is basically the relative change of y( in validation sample) on y\_hat( from LS predictor) with the expectation of the shrinkage is strictly less than one. The main criteria for prediction performance in this paper is PMSE( prediction mean square error) and prediction performance can be improved by reducing the dimension of predictive factors with formula proof.

In another paper of Copas, the reason of shrinkage occurrence is explained by ‘There is a sense in which the estimated model fits the data too well - shrinkage occurs because any unusual random features of the original data will be reflected in the predictions but not be replicated in a set of independent observations’. This may be because of the uniqueness of each set of data. The values of response variable for the new patients will be closer to the overall mean than we would expect from an uncritical application of least squares or maximum likelihood.

If the same set of data is used both to select the covariate and to fit the model, then the values of the coefficients will be biased and the shrinkage of the predictor will be even more marked. This is because that if a regression coefficient is by chance overestimated (in absolute value) then it will be more likely to be selected than if it happened to be underestimated, hence the covariates which end up being selected for the model are likely to have larger coefficients. This means that the LS predictor will give too wide a range of values, high values of y\_hat (prediction) will be too high and low values of y\_hat will be too small. This is why predictor selection methods would affect the shrinkage factor for preventing from over-fitting.