

5 Conclusions

In this thesis we study several aspects of variable selection problem for linear regressions, non-linear models and data with the presence of outliers.

The contributions of this work are a comparison of two of the most applied model selection techniques in the literature; a new methodology for variable selection for STR and STAR models; and a new method combining LASSO regularization and CVaR risk measure, interesting for data with outliers.

The new methodology for STR and STAR models presented satisfactory results on simulation exercises. The application to genetic data introduced us to a new field where databases can be of high dimension, but small samples. In this context, variable selection becomes of extreme importance for statistical analysis.

Chapter 4 presents first results on a new regularization method using a CVaR of “out-of-sample” errors as a second penalty term. The method has shown some benefits and some drawback. This is a first analysis of this approach and shall be developed.

We think we have made some progress in the field of variable selection, especially when models have more candidate variables than observations. We have some promising applications in genetic and energy data, which can be of very high dimension. The conclusions are detailed at the end of each chapter.

Variable selection and regularization methods are a very rich field of study, and this thesis opens some promising paths for future research.