

# Division of Statistics

## Master's Thesis Defense

**Lin Lu**

**Summer 2002**

### **A Comparison of Lasso and Bayesian Methods for Linear Models**

#### **ABSTRACT**

For a given data set with response variable  $y$  and a set of possible predictor variables,  $x_1, x_2, \dots, x_p$ , there are two common objectives of fitting a linear model of  $y$  as a function of the  $x$ 's. One objective is to quantify the nature of the relationship between  $y$  and the  $x$ 's, and the other is to use the relationship to predict the value of  $y$  given a new set of values of  $x$ 's. A valid model is one where the coefficients, accurately describe the nature of the relationship between the predictors and response variable; for example, the signs and relative magnitudes of the coefficients are correct.

Ordinary Least Squares (OLS) is a popular method of estimation. But OLS estimates can often be imprecise when the sample size is relatively small or multicollinearity is severe. Various methods of model selection have been developed to select a subset of the predictor variables ( $x$ 's) for use in a single model. But some selection methods, such as subset selection, can be quite unstable (Breiman 1996).

The least absolute shrinkage and selection operator, Lasso (Tibshirani 1996) is a shrinkage procedure similar to ridge regression but with the advantage that some covariates are removed from the model. Lasso is trying to select a single model to achieve both the above objectives.

The limitation of single model is ignoring model uncertainty. So model averaging may be considered. Bayesian model averaging is a model averaging method, which is aimed at the prediction case by averaging the possible models.

The thesis compared prediction ability and model validity among Lasso, the Bayesian method and OLS. The final results show that OLS performed poorly compared to the Bayesian method and Lasso when sample size is relative small to model complexity and multicollinearity is severe. The Bayesian method and Lasso are equivalent in terms of prediction accuracy. The Bayesian method performed better than Lasso, and Lasso performed better than OLS in terms of model validity.