Bayesian hierarchical regression via subspace embeddings*

Jonathan Rathjens†

5 December 2014

Abstract

Subspace embedding is a well known method of data reduction preserving the essential information. Its applicability in a Bayesian linear regression framework has already been proven. The posterior distribution of the coefficients estimated from the original data is approximated by that from the so-called sketch up to a small controlled error.

As a generalization, some hierarchical regression models are analyzed in this thesis. Apart from the regression coefficients, hyperparameters are to be estimated, each with comparison between both original and sketched data.

Simulation studies suggest a good approximation of both the regression coefficients and hyperparameters for all linear regression models with normal likelihood and various priors considered. For a generalized model with a logit-based link function, the approximation seems worse.

For a normal likelihood, a bounded distance between the two posterior distributions from the original and the sketched data can be derived from the non-hierarchical results, although without exact quantification. For the marginal distribution of the regression coefficients with normal priors, the previous result is reproduced exactly. If a location hyperparameter is the expected value of a regression coefficient, the same bound holds for it. A non-linear link function raises certain problems if a linear embedding is applied.

*Master's thesis, TU Dortmund University, Department of Statistics

†jonathan.rathjens@tu-dortmund.de