
# (Bayesian) Regression for Big Data using Random Projections

Leo Geppert
(TU Dortmund)

January 11, 2017, 16:00 s.t.
Department of Statistics, Ludwigstr. 33, Room 144, Munich

This talk is on the topic of random projections applied as a data reduction technique for Bayesian Regression. We show sufficient conditions under which the entire $p$-dimensional distribution is approximately preserved under random projections by reducing the number of data points from $n$ to $k \in O(\text{poly}(p/\varepsilon))$ in the case $n \gg p$. Under mild assumptions, we prove that evaluating a Gaussian likelihood function based on the projected data instead of the original data yields a $(1 + O(\varepsilon))$-approximation in terms of the $\ell_2$ Wasserstein distance. Our main result shows that the posterior distribution of Bayesian linear regression analyses is approximated up to a small error depending only on an $\varepsilon$-fraction of its defining parameters. This holds when using arbitrary Gaussian priors or the degenerate case of uniform distributions over $\mathbb{R}^p$ for $\beta$. Our empirical evaluations involve different simulated settings of Bayesian linear regression and a real data example. Our experiments underline that the proposed method is able to recover the regression model up to small error while considerably reducing the total running time and the required memory. Some generalizations to more complex regression models (hierarchical models, $q$-normal distributions as prior or likelihood, Poisson regression models) are also examined. Most of the results can directly be applied to frequentist regression models as well.