



Nonparametric Bayesian Label Prediction on a Graph
J. Hartog

SUMMARY

This dissertation is devoted to nonparametric Bayesian label prediction on a graph. Label prediction is a problem within the area of statistical learning. Typically, we want to predict a label, i.e. the outcome of some quantitative or categorical measurement, based on some features of the available data. A training set of observed, possibly noisy, labels and the corresponding features is available to base our inference on.

Label prediction problems on graphs arise in a variety of applications, for instance in machine learning, in the prediction of the biological function of a protein in a protein-protein interaction graph, in image analysis, and in the prediction of brand preference in social networks.

Prime examples are problems in which the graph is given by the application context. In our setting, we think of the labels as a stochastic process indexed by the graph vertex set. The data are noisy observations of the vertex labels. In case of regression on the graph, this process is real-valued. For classification problems, the process can be binary-valued or discrete-valued in general. Our main goal is to predict the missing labels correctly. The underlying idea is that the location of a given vertex in the graph and the information of its neighbors should have predictive power for the label of the vertex of interest.

First, we consider a hierarchical Bayesian approach with a randomly scaled Gaussian prior. A benefit of the hierarchical Bayesian procedure is that it is capable, when properly designed, to automatically determine the correct value for the scale parameter. The prior can be interpreted as a series expansion of the function of interest using the eigenvectors of the graph Laplacian matrix as a basis. In order to alleviate the computational burden of taking into account all the eigenvectors, we truncate the series at a random point to make our method applicable for very large graphs. For the full Bayesian approach, we sample from the joint posterior of the function of interest, regularization parameter and truncation level. Alternatively, we use empirical Bayes methods to select the regularization parameter and truncation level. After fixing these parameters, we sample from the posterior distribution of the function of interest. Finally, we use variational inference to approximate the posterior distribution. A major benefit of variational inference is that it allows for stochastic optimization to make it applicable to large data.