

Probabilistic Models in Bioinformatics

Sepp Hochreiter

JKU Linz

Abstract

I will show some applications of latent variable models in bioinformatics. These applications include microarray summarization, copy number detection, identification of differentially expressed genes in RNA-seq data, and biclustering of transcriptomic data. Latent variable models allow decomposing the data variation into a noise part and a signal part captured by the hidden variables. The posterior variance of the hidden variables served as informative/non-informative (I/NI) call for a single measurement similar to a signal-to-noise ratio. I/NI calls excelled as filter methods for gene expression data. For some applications we assumed a sparse prior for the latent variable like a Laplacian in order to model rare events like rare copy numbers or rare mutations. However, sparse priors on the hidden variables lead to an intractable likelihood or intractable parameter posterior. Hence, we applied variational methods to parameter selection in particular variational expectation maximization. I will present some examples, where latent variable models led to new biomedical discoveries in drug design projects and clinical studies.