

Protein Classification with Hidden Variables

For using support vector machines (Fig. 2), a kernel function should be defined a priori. We propose a reasonable way of designing a kernel when objects are generated from latent variable models (e.g. HMM). First of all, a joint kernel is designed for complete variables (i.e. both visible

and hidden). Then the hidden variables are marginalized out to give a marginalized kernel for visible variables. Although this framework can be applied to any object, we particularly derive several marginalized kernels useful for biological sequences (Fig. 1).

Koji TSUDA
Computational Biology
Research Center
e-mail:
koji.tsuda@aist.go.jp
AIST Today Vol. 2, No.8
(2002) 15

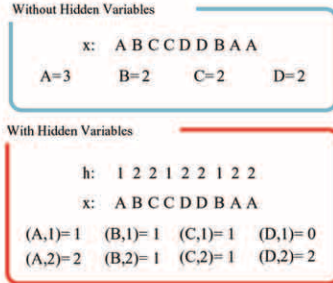


Fig.1 Feature Extraction with or without hidden variables
Dioxide

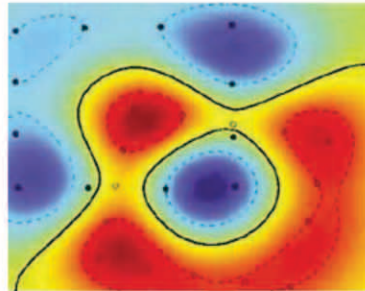


Fig.2 Partitioning a vector space with the support vector machines