Chapter 9
Sigma Tuning of Gaussian Kernels:
Detection of Ischemia from Magnetocardiograms

Long Han
Rensselaer Polytechnic Institute, USA

Mark J. Embrechts
Rensselaer Polytechnic Institute, USA

Boleslaw K. Szymanski
Rensselaer Polytechnic Institute, USA

Karsten Sternickel
Cardiomag Imaging, Inc., USA

Alexander Ross
Cardiomag Imaging, Inc., USA

ABSTRACT
This chapter introduces a novel Levenberg-Marquardt like second-order algorithm for tuning the Parzen window $\sigma$ in a Radial Basis Function (Gaussian) kernel. In this case, each attribute has its own sigma parameter associated with it. The values of the optimized $\sigma$ are then used as a gauge for variable selection. In this study, the Kernel Partial Least Squares (K-PLS) model is applied to several benchmark data sets in order to estimate the effectiveness of the second-order sigma tuning procedure for an RBF kernel. The variable subset selection method based on these sigma values is then compared with different feature selection procedures such as random forests and sensitivity analysis. The sigma-tuned RBF kernel model outperforms K-PLS and SVM models with a single sigma value. K-PLS models also compare favorably with Least Squares Support Vector Machines (LS-SVM), epsilon-insensitive Support Vector Regression and traditional PLS. The sigma tuning and variable selection procedure introduced in this chapter is applied to industrial magnetocardiogram data for the detection of ischemic heart disease from measurement of the magnetic field around the heart.

DOI: 10.4018/978-1-60960-551-3.ch009
BACKGROUND OF SIGMA TUNING

This chapter introduces a novel tuning mechanism for Gaussian or Radial Basis Function (RBF) kernels where each attribute (or feature) is characterized by its own Parzen window sigma. The kernel trick is frequently used in machine learning to transform the input domain into a feature domain where linear methods are then used to find an optimal solution to a regression or classification problem. Support Vector Machines (SVM), Kernel Principal Component Regression (K-PCR), Kernel Ridge Regression (K-RR), Kernel Partial Least Squares (K-PLS) are examples of techniques that apply kernels for machine learning and data mining. There are many different possible kernels, but the RBF (Gaussian) kernel is one of the most popular ones. Equation (1) represents a single element in the RBF kernel,

$$k(i, j) = e^{-\frac{(x_i - x_j)^2}{2\sigma^2}}$$  \hspace{1cm} (1)

where $x_i$ and $x_j$ denote two sample data. Traditionally, most machine learning approaches use a single value $\sigma$ in the RBF kernel (as indicated in the equation above), which then needs to be tuned on a validation or tuning data set. In this paper each attribute is associated with a different $\sigma$ value which is then tuned based on a validation data set with the aim to achieve a prediction performance that is an improvement over the one achieved by the RBF kernels with a single $\sigma$. The expression for a single RBF kernel entry becomes,

$$k(i, j) = \prod_{l=1}^{m} e^{-\frac{(x_{il} - x_{jl})^2}{2\sigma_{l}^2}}$$  \hspace{1cm} (2)

where $m$ is the number of attributes in the sample data. There are several advantages of using an automated tuning algorithm for a vector of $\sigma$ rather than selecting a single scalar variable:

- Manual tuning for multiple $\sigma$-values is a tedious procedure;
- The same automated procedure applies to most machine learning methods that use an RBF kernel;
- The values of the optimized $\sigma$ can be used as a gauge for variable selection (Specht, 1990).

LITERATURE OVERVIEW

Automated tuning of the kernel parameters is an important problem, it could be used in all different scientific applications: such as image classification (Guo, 2008; Claude, 2010) and time series data forecasting (He, 2008; Rubio, 2010), etc. A number of researchers have proposed algorithms for solving it, especially in the context of SVMs. Related work includes Grandvalet et al. (Grandvalet, 2002), which introduced an algorithm for automatic relevance determination of input variables in SVMs. Relevance is measured by scale factors defining the input space metric. The metric is automatically tuned by the minimization of the standard SVM empirical risk, where scale factors are added to the usual set of parameters defining the classifier. Cristianini et al. (Cristianini, 1998) applied an iterative optimization scheme