Exercises for
Pattern Recognition
Peter Fischer, Shiyang Hu
Assignment 9, 16./19.12.2013

General Information:

Exercises (1 SWS):	Tue 12:15 – 13:45 (0.154-115) and Fri 08:15 – 09:45 (0.151-115)
Certificate:	Oral exam at the end of the semester
Contact:	peter.fischer@fau.de, shiyang.hu@fau.de

Support Vector Regression

Exercise 1 In the lecture, you learn how an SVM can be used for classification. In this exercise, we consider *Support Vector Regression* (SVR). Let $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}, \boldsymbol{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R}$ be a set of observations. The task for regression is to predict y_i from $\boldsymbol{x}_i$ according to the linear regression function:

$$F(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{x} + \alpha_0, \tag{1}$$

for a weight vector $\boldsymbol{\alpha} \in \mathbb{R}^d$ and the bias $\alpha_0 \in \mathbb{R}$. The intuition behind SVR is to penalize only deviations that are larger than ϵ.

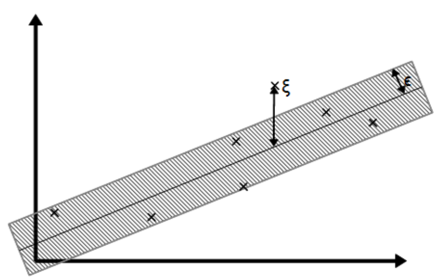

Figure 1: ϵ-tube of the SVR

The primal optimization problem for SVR is given by the following inequality-constraint minimization:

$$\begin{aligned}
\boldsymbol{\alpha}^* &= \operatorname*{argmin}_{\boldsymbol{\alpha}} \frac{1}{2}||\boldsymbol{\alpha}||^2 + C \sum_i (\xi_i + \hat{\xi}_i) \text{ , s.t.} \\
y_i &\leq \left(\boldsymbol{\alpha}^T \boldsymbol{x}_i + \alpha_0\right) + \epsilon + \xi_i \\
y_i &\geq \left(\boldsymbol{\alpha}^T \boldsymbol{x}_i + \alpha_0\right) - \epsilon - \hat{\xi}_i \\
\xi_i, \hat{\xi}_i &\geq 0
\end{aligned}$$

Here, ξ_i, $\hat{\xi}_i$ are slack variables (see also SVM classification) and ϵ specifies uncertainty of the regression function.

(a) Write down the Lagrangian L of the primal optimization problem using Lagrange multipliers λ_i, $\hat{\lambda}_i$, μ_i, $\hat{\mu}_i$.
Hint: bring the constraints to the standard form $f_i(\boldsymbol{x}) \leq 0$

(b) Write down the Karush-Kuhn-Tucker (KKT) conditions for the primal optimization problem given above.

(c) Derive the dual optimization problem. To derive the dual optimization problem, you have to eliminate $\boldsymbol{\alpha}$, $\boldsymbol{\xi}$, and $\hat{\boldsymbol{\xi}}$ from L using the gradient of L. Preliminary solution:

$$\begin{aligned} L\left(\boldsymbol{\alpha}, \alpha_0, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}, \boldsymbol{\mu}, \hat{\boldsymbol{\mu}}\right) =& \frac{1}{2}||\boldsymbol{\alpha}||^2 + C\sum_i (\xi_i + \hat{\xi}_i) + \sum_i \left(-\mu_i \xi_i - \hat{\mu}_i \hat{\xi}_i\right) + \\ & \sum_i \lambda_i \left(y_i - \boldsymbol{\alpha}^T \boldsymbol{x}_i - \alpha_0 - \epsilon - \xi_i\right) + \\ & \sum_i \hat{\lambda}_i \left(-y_i + \boldsymbol{\alpha}^T \boldsymbol{x}_i + \alpha_0 - \epsilon - \hat{\xi}_i\right) \end{aligned}$$

(d) Which property must be fulfilled for *support vectors* in SVR?
Hint: replace $\boldsymbol{\alpha}$ in Equation (1).