Machine learning 2
Exercise sheet 5

FLEISCHMANN Kay, Matrnr: 352247
ROHRMANN Till, Matrnr: 343756

May 14, 2013

1 One-class-SVM: Theory

(a) Derive the dual program for the one-class SVM.

Primal form

The primal form of the one-class SVM has the following form:

$$\min_{\boldsymbol{\mu},r,\boldsymbol{\xi}} r^2 + C\sum_{i=1}^{N}\xi_i$$

such that

$$\begin{aligned}\|\phi(x_i) - \boldsymbol{\mu}\|^2 &\leq r^2 + \xi_i \\ \xi_i &\geq 0\end{aligned}$$

for $i = 1, \dots, n$. Using Lagrange multipliers gives us the unconstrained form:

$$\min_{\boldsymbol{\mu},r,\boldsymbol{\xi}} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}\geq 0} \underbrace{\left\{ r^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\alpha_i\left(\|\phi(x_i) - \boldsymbol{\mu}\|^2 - r^2 - \xi_i\right) - \sum_{i=1}^{N}\beta_i\xi_i \right\}}_{L(\boldsymbol{\mu},r,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\beta})}$$

The dual optimization problem is now given by

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}\geq 0} g(\boldsymbol{\alpha},\boldsymbol{\beta})$$

with g being defined by

$$g(\boldsymbol{\alpha},\boldsymbol{\beta}) = \min_{\boldsymbol{\mu},r,\boldsymbol{\xi}} L(\boldsymbol{\mu}, r, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{1}$$

To compute the minimum of L w.r.t. $\boldsymbol{\mu}, r$ and $\boldsymbol{\xi}$ we take the partial derivative and set it afterwards to zero.

$$\begin{aligned}\nabla_{\boldsymbol{\mu}} L &= \nabla_{\boldsymbol{\mu}}\left(\sum_{i=1}^{N}\alpha_i\left(\phi(x_i) - \boldsymbol{\mu}\right)^T\left(\phi(x_i) - \boldsymbol{\mu}\right)\right) \\ &= \sum_{i=1}^{N}\alpha_i\left(2\boldsymbol{\mu} - 2\phi(x_i)\right) \qquad (2) \\ \frac{\partial L}{\partial r} &= 2r - 2\sum_{i=1}^{N}\alpha_i r \qquad (3) \\ \frac{\partial L}{\partial \xi_j} &= C - \alpha_j - \beta_j \qquad (4)\end{aligned}$$

Setting equations (2),(3) and (4) to 0 we obtain

$$\begin{aligned}\boldsymbol{\mu}\sum_{i=1}^{N}\alpha_i &= \sum_{i=1}^{N}\alpha_i\phi(x_i) \qquad (5) \\ (1 - \sum_{i=1}^{N}\alpha_i)r &= 0 \qquad (6) \\ C &= \alpha_i + \beta_i \qquad (7)\end{aligned}$$

Assuming that we have at least 2 distinct data points, we know that $r > 0$ holds. Thus equation (6) gives us

$$\sum_{i=1}^{N} \alpha_i = 1 \tag{8}$$

and thus equation (5) can be expressed by

$$\boldsymbol{\mu} = \sum_{i=1}^{N} \alpha_i \phi(x_i) \tag{9}$$

This equation says that one can express the optimal solution for $\boldsymbol{\mu}$ as a linear combination of the data points in feature space. Plugging equations (7) and (9) into equation (1) gives us

$$\begin{aligned} g(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= r^2 + \sum_{i=1}^{N}(\alpha_i + \beta_i)\xi_i + \sum_{i=1}^{N} \alpha_i \left(\|\phi(x_i) - \sum_{i=1}^{N} \alpha_i \phi(x_i)\|^2 - r^2 - \xi_i \right) - \sum_{i=1}^{N} \beta_i \xi_i \\ &= r^2 - r^2 \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \alpha_i \left(\phi(x_i) - \sum_{j=1}^{N} \alpha_j \phi(x_j) \right)^T \left(\phi(x_i) - \sum_{j=1}^{N} \alpha_j \phi(x_j) \right) \end{aligned}$$

Using equation (8) gives us

$$g(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \phi(x_i) - \sum_{i,j=1}^{N} \alpha_i \alpha_j \phi(x_i)^T \phi(x_j)$$

with the additional constraints

$$\begin{aligned} \sum_{i=1}^{N} \alpha_i &= 1 \\ C = \alpha_i + \beta_i &\Rightarrow 0 \leq \alpha_i \leq C \end{aligned}$$

Assuming we have a kernel function k expressing the inner product $\phi(x)^T \phi(y) = k(x, y)$ we finally end up at the final formulation:

$$\max_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{N} \alpha_i k(x_i, x_i) - \sum_{i,j=1}^{N} \alpha_i \alpha_j k(x_i, x_j) \right\} \tag{10}$$

subject to

$$\begin{aligned} \sum_{i=1}^{N} \alpha_i &= 1 \\ 0 \leq \alpha_i &\leq C \text{ with } i = 1, \ldots, n \end{aligned}$$

(b) Show that the dual problem is a linearly constrained quadratic problem.

Setting $(\boldsymbol{b})_i = k(x_i, x_i)$ and $(A)_{i,j} = -k(x_i, x_j)$ we can reformulate equation (10) in its matrix/vector notation

$$(10) \quad = \quad \max_{\boldsymbol{\alpha}} \boldsymbol{\alpha}^T A \boldsymbol{\alpha} + \boldsymbol{b}^T \boldsymbol{\alpha}$$

Furthermore by setting $v = 1$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, $l_i = 0$ and $m_i = C$ for $i = 1, \ldots, n$ we can rewrite the constraints:

$$\begin{aligned} \sum_{i=1}^{N} \alpha_i = 1 \quad &\Leftrightarrow \quad \boldsymbol{u}^T \boldsymbol{\alpha} = v \\ 0 \leq \alpha_i \leq C \quad &\Leftrightarrow \quad l_i \leq \alpha_i \leq m_i \end{aligned}$$