

# Lab 1 Machine Learning

Anton Stråhle & Jan Alexandersson

10 februari 2020

## Dual formulation

The minimization of the objective can be written as

$$\min_{\alpha_1, \dots, \alpha_n} \frac{1}{2} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}^T \begin{bmatrix} t_1 t_1 \kappa(\vec{x}_1, \vec{x}_2) & \dots & t_1 t_n \kappa(\vec{x}_1, \vec{x}_n) \\ \vdots & \ddots & \vdots \\ t_n t_1 \kappa(\vec{x}_n, \vec{x}_1) & \dots & t_n t_n \kappa(\vec{x}_n, \vec{x}_n) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}^T \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

with the corresponding constraints

$$\begin{bmatrix} -1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & & -1 \\ 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \preceq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ C \\ \vdots \\ C \end{bmatrix}.$$

We solve this using the QP function in the CVXOPT package.

## Assignment 1

- An optimal solution can not be found using a linear kernel if the classes are not linearly separable.
- An optimal solution can not be found if the data is not separable with the polynomial degree used. For example, degree 3 may work but degree 2 may not, then data is separable with the degree 3.
- When using a radial kernel, an optimal solution can not be found if the data is not separable at all.

## **Assignment 2**

## **Assignment 3**

When using a polynomial kernel a higher degree or will classify more precise according to the training set which will increase the variance but decreasing the bias. Similarly, when using a Radial Basis Function kernel a lower sigma will increase variance and decrease bias. This may lead to overfitting.

## **Assignment 4**

If  $C$  is small, then there will be fewer support vectors and hence the resulting classifier will have low bias but high variance. This is because a low value of  $C$  does not allow for the margins to be as violated as a high value.

## **Assignment 5**

If the dataset is noisy it is better to have low/high slack??? Need to think more about this.