Machine Learning Course 2024 Spring: Homework 2

March 31, 2024

1 Problem 1

Recall the primal problem of hard-margin linear SVM is

$$\begin{aligned} \min_{\boldsymbol{w},b} \quad & \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \ldots, m \end{aligned} \tag{1.1}$$

and soft-margin linear SVM is

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} + C\sum_{i=1}^{m} \xi_i \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{1.2}$$

The linear classifiers Eq.(1.1) and Eq.(1.2) can be induced by solving the dual of this primal problem to perform linear classification.

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \alpha_i y_i \left(\boldsymbol{x}_i^\top \boldsymbol{x}\right) + b \tag{1.3}$$

Moreover, we can replace $\boldsymbol{x}_i^\top \boldsymbol{x}$ with a kernel $\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}\right)$ to achieve a non-linear classifier.

In Figure 1, there are 6 different SVMs with different patterns of decision boundaries. The training data is labeled as $y \in \{-1, 1\}$, represented as the shape of circles and squares respectively. Support vectors are drawn in SOLID circles. Label each plot in Figure 1 with the letter of the optimization problem below and explain WHY you pick the figure for a given kernel. (Note that one of the plots does not match to anything.)

1. A soft-margin linear SVM with $C = 0.02$.
2. A soft-margin linear SVM with $C = 20$.

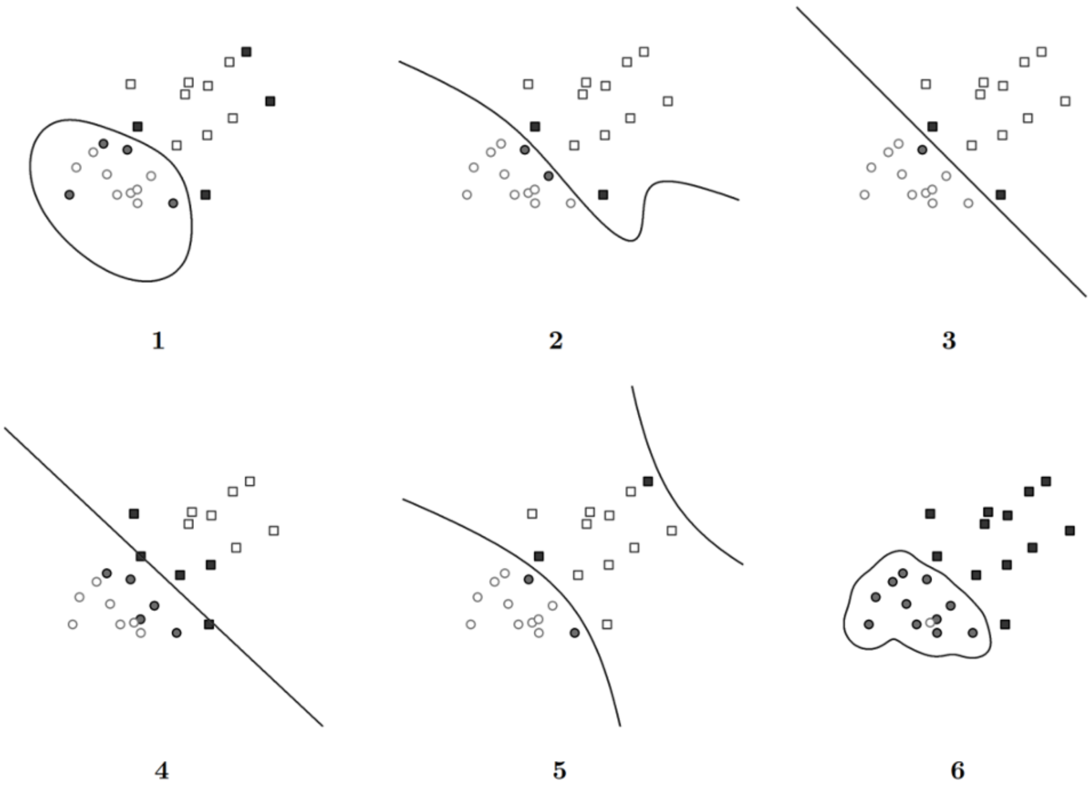

Figure 1: SVM Decision Boundaries

3. A hard-margin kernel SVM with $\kappa(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^\top \boldsymbol{v} + \left(\boldsymbol{u}^\top \boldsymbol{v}\right)^2$.
4. A hard-margin kernel SVM with $\kappa(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-5\|\boldsymbol{u} - \boldsymbol{v}\|^2\right)$.
5. A hard-margin kernel SVM with $\kappa(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-\frac{1}{5}\|\boldsymbol{u} - \boldsymbol{v}\|^2\right)$.

2 Problem 2

Consider the following convex optimization problem:

$$\begin{aligned} \min_{x_1, \ldots, x_n} \quad & -\sum_{i=1}^{n} \log(\alpha_i + x_i) \\ \text{s.t.} \quad & x_i \geq 0, \sum_{i=1}^{n} x_i = 1, \end{aligned} \tag{2.1}$$

where $\alpha_i > 0$.

1. Write the KKT conditions for the problem.

2. Find the optimal solution of the problem.

3 Problem 3

Kernel functions implicitly define some mapping function $\phi(\cdot)$ that transforms an input instance $\boldsymbol{x} \in \mathbb{R}^d$ to a high dimensional space Q by the form of dot product: $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j)$.

1. Assume that we are using an RBF kernel function $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\frac{1}{2}||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)$, where an implicit unknown mapping function $\phi(\cdot)$ exists. Prove that for any two input instances, the squared Euclidean distance of their corresponding points in the feature space Q is less than 2, i.e., $||\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j)|| < 2$.

2. Consider a kernel function on $\mathbb{R}^d$: $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j + c)^N$, where c is any real number, d, N are any positive integers. Try analyzing when function κ is/is not a kernel function, and state your reasons.