Machine Learning Worksheet 08

Soft-margin SVM and Kernels

1 Soft-margin SVM

Problem 1: Assume that we have a linearly separable dataset $\mathcal{D}$, on which a soft-margin SVM is fitted. Is it guaranteed that all training samples in $\mathcal{D}$ will be assigned the correct label by the fitted model? Explain your answer.

Problem 2: Why do we need to ensure that $C > 0$ in the slack variable formulation of soft-margin SVM? What would happen is this was not the case?

2 Kernels

Problem 3: Show that for $c \geq 0$ and $d \in \mathbb{N}^+$ the function

$$K(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}^T \boldsymbol{y} + c)^d$$

is a valid kernel.

3 Gaussian kernel

Problem 4: Let us define a feature transformation $\boldsymbol{\phi_n} : \mathbb{R} \to \mathbb{R}^n$ as follows:

$$\boldsymbol{\phi_n}(x) = \left\{ e^{-x^2/2\sigma^2}, e^{-x^2/2\sigma^2} \frac{x}{\sigma}, \frac{e^{-x^2/2\sigma^2} \left(\frac{x}{\sigma}\right)^2}{\sqrt{2}}, \dots, \frac{e^{-x^2/2\sigma^2} \left(\frac{x}{\sigma}\right)^i}{\sqrt{i!}}, \dots, \frac{e^{-x^2/2\sigma^2} \left(\frac{x}{\sigma}\right)^n}{\sqrt{n!}} \right\}$$

Suppose we let $n \to \infty$ and define a new feature transformation:

$$\boldsymbol{\phi_\infty}(x) = \left\{ e^{-x^2/2\sigma^2}, e^{-x^2/2\sigma^2} \frac{x}{\sigma}, \frac{e^{-x^2/2\sigma^2} \left(\frac{x}{\sigma}\right)^2}{\sqrt{2}}, \dots, \frac{e^{-x^2/2\sigma^2} \left(\frac{x}{\sigma}\right)^i}{\sqrt{i!}}, \dots \right\}$$

Can we directly apply this feature transformation to data? Explain why or why not!

Problem 5: From the lecture, we know that we can express a linear classifier using only inner products of input vectors in the transformed feature space. It would be great if we could somehow use the feature space obtained by the feature transformation $\boldsymbol{\phi_\infty}$. However, to do this, we must be able to compute the inner product of samples in this infinite feature space. We define the inner product between two *infinite* vectors $\boldsymbol{\phi}_{\infty}(x)$ and $\boldsymbol{\phi}_{\infty}(y)$ as the infinite sum given in the following equation:

$$K(x, y) = \sum_{i=0}^{\infty} \boldsymbol{\phi}_{\boldsymbol{\infty},\boldsymbol{i}}(x)\, \boldsymbol{\phi}_{\boldsymbol{\infty},\boldsymbol{i}}(y)$$

What is the explicit form of $K(x, y)$? (Hint: Think of the Taylor series of e^x.) With such a high dimensional feature space, should we be concerned about overfitting?

Problem 6: Can any *finite* set of points be linearly separated in the feature space defined by $\boldsymbol{\phi}_{\infty}$ if σ can be chosen freely?

4 Kernelized k-nearest neighbors

To classify the point $\boldsymbol{x}$ the k-nearest neighbors finds the k training samples $\mathcal{N} = \{\boldsymbol{x}^{(s_1)}, \boldsymbol{x}^{(s_2)}, \ldots, \boldsymbol{x}^{(s_k)}\}$ that have the shortest distance $||\boldsymbol{x} - \boldsymbol{x}^{(s_i)}||_2$ to $\boldsymbol{x}$. Then the label that is mostly represented in the neighbor set $\mathcal{N}$ is assigned to $\boldsymbol{x}$.

Problem 7: Formulate the k-nearest neighbors algorithm in feature space by introducing the feature map $\boldsymbol{\phi}(\boldsymbol{x})$. Then rewrite the k-nearest neighbors algorithm so that it only depends on the scalar product in feature space $K(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{\phi}(\boldsymbol{y})$.