hw3

Wenhua Bao 2512664
Xinrui Chen 2503398

August 2, 2020

1 Task2

1.0.1 a

1. SVM is the linear classifier with the largest spacing defined on the feature space . 2.Advantages: The classifier only depends on a few data points. Both the original SVM formulation (primal) as well as the derived dual formulation are quadratic programming problems (quadratic cost, linear constraints), which have unique solutions that can be computed efficiently

1.0.2 b

$$\begin{aligned} &\arg\min_{w,b} \frac{1}{2}\|w\|^2 \\ &\text{s.t. } y_i\left(w^\top x_i + b\right) - 1 \geqslant 0 \quad \forall i \end{aligned}$$

1.0.3 c

In reality the modes have noise to make models not linear separable, slack variables make the model can tolerant noise and become linear separable.

$$\begin{aligned} &\arg min\tfrac{1}{2}\|w\|^2 + C\textstyle\sum_{i=1}^N \zeta_i \\ &\text{sit. } y_i\left(w^\top x_i + b\right) \geqslant 1 - s_i \quad (i = 1, 2, \ldots N) \\ &\zeta_i \geqslant 0 \end{aligned}$$

1.0.4 d

$$\begin{gathered} L(\omega, b, \alpha) = \tfrac{1}{2}\|w\|^2 - \alpha_i \textstyle\sum_{i=1}^N (y_i\left(w^\top x_i + b\right) - 1\Big) \\ = \tfrac{1}{2}\|w\|^2 - \textstyle\sum_{i=1}^N \alpha_i y_i w^\top x_i - \sum_{i=1}^N \alpha_i y_i b + \sum_{i=1}^N \alpha_i \\ \tfrac{\partial L}{\partial w} = w - \textstyle\sum \alpha_i y_i x_i = 0 \\ w = \textstyle\sum \alpha_i y_i x_i \\ \tfrac{\partial L}{\partial b} = -\textstyle\sum \alpha_i y_i = 0 \\ \textstyle\sum \alpha_i y_i = 0 \end{gathered}$$

$$
\begin{aligned}
L(w, b, \alpha) &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(x_j^\top x_i\right) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i a_j y_i y_j \left(x_j^\top x_i\right) + \sum_{i=1}^{N} a_i \\
&= \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(x_j^\top x_i\right)
\end{aligned}
$$

result

$$
\begin{aligned}
\min \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left(x_i^\top x_j\right) \\
\text{s.t. } \alpha_i \geqslant 0 \\
\sum_{i=1}^{N} \alpha_i y_i = 0
\end{aligned}
$$

1.0.5 e

we can skip the derivation

1.0.6 f

1.Replace every occurrence of a scalar product between features with a kernel function

$$
K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \phi\left(\mathbf{x}_i\right)^\top \phi\left(\mathbf{x}_j\right)
$$

2.A kernel $K\left(\mathbf{x}_i, \mathbf{x}_j\right)$that allows us to compute the scalar product without making the mapping explicit

1.0.7 g

2 Task3

2.1 a