MLSM 2021 - Quiz 4

02/11/2021 - 10/11/2021 at 23:59

Questions based on Lecture 6 and 7

(1) (1.0 pt.) In the soft-margin SVM without considering the bias term we have the constraints: $y_i \mathbf{w}^T \mathbf{x}_i \geq 1 - \xi_i$ for all $i = 1, \ldots, m$. Assume that the $\mathbf{w}$ is the optimal solution of SVM problem. If a pair, $(y_i, \mathbf{x}_i)$, of a training example is not correctly classified which of these statement is true?

(1) $\xi_i = 0$
(2) $\xi_i < 0$
(3) $\xi_i \geq 1$
(4) $0 < \xi_i \leq 1$

(2) (1.0 pt.)

Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be two kernel matrices. They are positive semi-definite and symmetric, with the same size. Which of these three operations provide a correct kernel matrix in the general case?

(1) The linear combination of $\mathbf{K}_1$ and $\mathbf{K}_2$ with any real weights.
(2) The itemwise (pointwise, Hadamard) product of $\mathbf{K}_1$ and $\mathbf{K}_2$.
(3) The matrix product of $\mathbf{K}_1$ and $\mathbf{K}_2$.

(3) (1.0 pt.) Let the feature vector of a polynomial kernel, $\kappa_{pol}(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z} + c)^q$, be represented explicitly. What is the dimension of the explicit feature vector? Assume that the degree of the polynomial is 4, and the polynomial kernel is defined on the vector space of dimension 2.

(1) 6
(2) 15
(3) 20
(4) 10

(4) (2.0 pt.)

In this question the behavior of two algorithms developed to solve the Support Vector Machine problem are compared. To this end, load the breast cancer dataset from the sklearn package,

`from sklearn.datasets import load_breast_cancer`.

Scale each of the input variables to have the maximum absolute value equal to 1. For example, it can be implemented by

```
input numpy as np

## X is the input matrix
mdata,ndim = X.shape
X/= np.outer(np.ones(mdata),np.max(np.abs(X),0))
```

Compute the normal vector of the hyperplane $\mathbf{w}$ by the primal algorithm given by the Slide "Stochastic gradient descent algorithm for soft-margin SVM ". Process all examples in the given fix order as they are appearing in the data set. Repeat the processing of all data examples 10 times. The parameters defined on the slide are set to these values: penalty weight λ is 0.01, and the step size η is fixed and it equal to 0.1. The data set is not split into training and test, the entire set is used in the learning.

On the same data set also apply the "Stochastic Dual Coordinate Ascent for SVM" algorithm. In this case also process the examples in the given fix order, and repeat that 10 times similarly to the primal algorithm. The kernel is assumed to be linear, and the C constant is equal to 1000. Based on the solution vector $\boldsymbol{\alpha}$ another estimation of the normal vector $\mathbf{w}$ can be computed as well, see the Slide "Dual Soft-Margin SVM".

What is the Pearson correlation coefficient between the $\mathbf{w}$ computed in the primal algorithm, and that $\mathbf{w}$ which is provided by the dual algorithm? Round the two numbers up to 2 decimals, and take the closest case.

(1) 0.72
(2) 0.80
(3) 0.92
(4) 0.99