EN4573 Assignment 02: Kernel Methods and Maximum Margin Classifiers

M. T. U. Sampath K. Perera

March 29, 2024

1 kernel Methods

1. Suppose that input space to feature space mapping (projection) is given by the following function $\Phi : \boldsymbol{x} = (x_1,\ x_2) \to \Phi(\boldsymbol{x}) = \left(x_1^2,\ x_2^2,\ \sqrt{2}\, x_1 x_2\right) \in \mathbb{R}^3$. Show that the kernel function provided above for a two-dimensional input space is $k(\boldsymbol{x}, \boldsymbol{z}) = \langle \boldsymbol{x}, \boldsymbol{z} \rangle^2$. Here, $\boldsymbol{x} = (x_1, x_2)$, $\boldsymbol{z} = (z_1, z_2)$ and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ inner product between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

2. Consider the kernel $k = (\boldsymbol{x}^T \boldsymbol{z})^2$. For the following dataset, determine the kernel matrix (gram matrix). Is this a valid kernel for this data set? justify your answer.

Sample index	Data sample ($\boldsymbol{x} = (x_1, x_2)$)	Feature 1 (x_1)	Feature 2 (x_2)
1	$\boldsymbol{x}_1$	2	3
2	$\boldsymbol{x}_2$	1	2
3	$\boldsymbol{x}_3$	2	4

Note:- The kernel matrix (gram matrix) is given by

$$
\boldsymbol{G} = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & k(\boldsymbol{x}_1, \boldsymbol{x}_2) & \cdots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ k(\boldsymbol{x}_2, \boldsymbol{x}_1) & k(\boldsymbol{x}_2, \boldsymbol{x}_2) & \cdots & k(\boldsymbol{x}_2, \boldsymbol{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & k(\boldsymbol{x}_N, \boldsymbol{x}_2) & \cdots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{bmatrix}
$$

2 Maximum margin classifiers/SVM

1. The following equation represents the objective function of the Support Vector Machine (SVM) algorithm, which aims to minimize the margin of separation between classes while penalizing misclassifications. The parameter C, known as the regularization parameter, balances the trade-off between maximizing the margin and minimizing classification errors.

$$
\min_{\mathbf{w}, w_0} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \max\{0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i + w_0)\}. \tag{1}
$$

Modified version of the above equation is given below

$$\min_{\mathbf{w},w_0} \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{N}\sum_{j=1}^{k} c_j \left(\max\left(0, 1 - y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right)\right)\right). \tag{2}$$

Here, separate regularization parameter is applied for each class as given below

$$c_j = \begin{cases} c_0 & \text{if } y_i = -1, \\ c_1 & \text{if } y_i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

This link contains the both standard and modified version of the SVM implementation. Download the code and run it either your own laptop or your own virtual environment (e.g., colab).

(a) Run the first segment of the code (section Question 2(a) in the python file) and write-down the coordinates of the centers.

(b) Run the second segment of the code (section Question 2(b)) and attach figure of the decision boundary to your report. Comment on the nature of the decision boundaries of non-weight (standard SVM, eq.(1)) and weighted (Modified SVM, eq.(2)) algorithms.

(c) Run the third segment of the code (section Question 2(c)) and attach figure of the decision boundary to your report. Here, why decision boundaries of non-weight SVM and weighted SVM algorithms are different ?

(d) Run the fourth segment of the code (section Question 2(d)) and attach figure of the decision boundary to your report. What is the impact of the ratio of c_0/c_1 to the decision boundary of the weighted SVM algorithm.

(e) Run the fifth segment of the code (section Question 2(e)) and attach figure of the decision boundary to your report. What are your observations on generated data compared to section 2(a). Comment of the nature of the decision boundaries of two algorithms.

(f) Run the sixth segment of the code (section Question 2(f)) and attach figure of the decision boundary to your report. What are your observations on generated data compared to the section 2(e). Comment of the nature of the decision boundaries of two algorithms.

(g) Based on your results in sections 2(e) and 2(f) will it be better to use weighted SVM algorithm for class unbalanced dataset? justify your answer.

2. Support Vector Machines (SVMs) do not provides probability estimates. Instead, for binary case, probabilities are calculated using the method give in reference [4]. Briefly summarize this method and mention any weaknesses associated with it.

Submission

- Upload a report and your codes as a zip file named as "EN4573_your_indexno_A02.zip". Include the index number and the name within the report as well.
- The interpretation of results and the discussion are important in the report.
- Pay careful attention to formatting such as font size, spacing, and margins.
- Include a title page with necessary information (e.g., title, author, date, index no).
- Use consistent and professional formatting throughout the document.
- Plagiarism will be checked and in cases of plagiarism, an extra penalty of 50% will be applied. In case of copying from each other, both parties involved will receive a grade of zero for the assignment. Academic integrity is of utmost importance, and any form of plagiarism[1] or cheating will not be tolerated.
- An extra penalty of 10% is applied for late submission.

References

[1] Scikit-learn Custom Kernel. https://scikit-learn.org/stable/auto_examples/svm/plot_custom_kernel.html. [Accessed 29-03-2024].

[2] Scikit-learn Support Vector Machines Documentation. https://scikit-learn.org/stable/modules/svm.html. [Accessed 29-03-2024].

[3] Andrew Ng. Stanford CS229 Lecture Notes. https://cs229.stanford.edu/notes2021fall/cs229-notes3.pdf, 2019. [Accessed 29-03-2024].

[4] John Platt and Nikos Karampatziakis. Probabilistic outputs for SVMs and comparisons to regularized likelihood methods. *Adv. Large Margin Classif*, 10, 2007.

[1] https://en.wikipedia.org/wiki/Plagiarism