CSE455/555 - Intro to Pattern Recognition Problem Set 2: Linear Discriminant Functions and Support Vector Machines

Due Date: Friday, March 11, 2022 11:59PM

In this problem set, you will train a support vector classifier and report performance using MNIST testing data set. http://yann.lecun.com/exdb/mnist/. You will train your model using the training data sets (”train-images-idx3-ubyte.gz” and ”train-labels-idx1-ubyte.gz”) and test the performance using the test data set (”t10k-images-idx3-ubyte.gz” and ”t10klabels-idx1-ubyte.gz”). Then you will derive the primal-dual relationship of the 1-norm softmargin classification problem, in which process you will demonstrate your understanding of several key concepts, such as maximal margin and support vector. The emphasis is on getting hands dirty with SVM and understanding the theory. There are many tutorials already on this problem, and you can refer the solution to problem 30 in Chapter 5 on page 275 of textbook “Pattern Classification” for a similar mathematical derivation involving the method of Lagrange multipliers and KKT condition. Note: there is no need to wait until we cover support vector machine to begin solving the problems.

1 Task-1

Write code to train a multi-class support vector classifier with dot-product kernel and 1-norm soft margin using the MNIST training data set. Then report the performance using MNIST test data set. There is a hyper-parameter that sets the trade-off between the margin and the training error — tune this hyper-parameter through cross-validation.

2 Task-2

Identify the Lagrange dual problem of the following primal problem:

Given features $(x_1,y_1),...,(x_N,y_N)$, where $y_1,...,y_N \in \{-1,1\}$. Minimize $w^T \cdot w + C \sum_{i=1}^{N} \xi_i$, the weighted sum between the squared length of the separating vector and the errors, where w is the separating vector, $w^T \cdot w$ is the dot product, and ξ_i is the error made by separating vector w on feature (x_i,y_i), subject to $y_i \cdot (w^T \cdot x_i) \geq 1 - \xi_i$ and $\xi_i \geq 0$ for $i = 1,...,N$. In other words, if the “normalized feature” $y_i x_i$ has a margin less than 1, $w^T \cdot (y_i x_i) \leq 1$, we add a slackness term to make it 1.

Point out what is the ”margin” in both the primal formulation and the dual formulation, what are the benefits of maximizing the margin. Characterize the support vectors. Point out the benefit of solving the dual problem instead of the primal problem.

3 Task-3[Optional]

Formulate the primal problem and derive the dual problem if there are multiple classes.

4 Submission

Submit your solutions as one ipynb file through UBlearn. You can use Google Colab:
https://colab.research.google.com/notebooks/intro.ipynb
https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c. The ipynb file should include your code, execution results, any explanations and answers to the questions. Use text cells to answer questions and add explanations. Markdown guide for text cells:
https://colab.research.google.com/notebooks/markdown_guide.ipynb#scrollTo=Lhfnlq1Surtk
https://colab.research.google.com/notebooks/basic_features_overview.ipynb#scrollTo=4hfV37gxpP_c
You can also add math to text cells using LaTeX. Just place the statement within a pair of $ signs. Please typeset your mathematics. Do not upload pictures of handwriting math formulas. Math typesetting help: https://www.codecogs.com/latex/eqneditor.php

5 Libraries

Basic libraries are allowed, such as gzip, pickle, math, numpy, scipy, matplotlib, etc.
You can also use Keras.datasets to import MNIST. Scikit-learn is
allowed for this assignment.

6 Rubric

Total: 10 points + 2 bonus points

Task-1: 2 points
Task-2: 8 points.
Task-3: 2 bonus points.

7 Acknowledgement

(3) that your use of SafeAssign will be without recourse against Blackboard Inc. and its affiliates.

8 Academic Integrity

Academic integrity is a fundamental university value. Any violation will be reported to the University and will result in penalties in grades.

Do not share your answers with other students. This is an individual assignment. You are not allowed to work in groups. Working in groups and submitting similar answers is considered a violation of academic integrity.

Do not plagiarize someone else's words, ideas, or data you find online. Always cite your sources.