Pattern Recognition: Homework 8

Due date: 2023.4.25

Problem 1

Introduction:

In this assignment, we will explore Laplacian Regularized Least Squares (LapRLS) for multi-class classification in a semi-supervised setting, where the number of classes is denoted as c. Our baseline model is Regularized Least Square (RLS), a linear regression model that includes an additional regularization term to control the complexity of the model. LapRLS is an extension of RLS that exploits the underlying manifold structure of the data.

The optimization objective for RLS is given by:

$$\min_{W\in\mathbb{R}^{l\times c}} \frac{1}{l}\mathrm{Tr}\big((Y-KW)^T(Y-KW)\big)+\gamma\mathrm{Tr}(W^TKW) \tag{1}$$

where K is an $l\times l$ Gram matrix of a pre-defined kernel over labeled points, $Y\in R^{l\times c}$ represents the corresponding one-hot label, and l is the number of labeled samples.

The optimization objective for LapRLS is expressed as:

$$\min_{W\in\mathbb{R}^{(l+u)\times c}} \frac{1}{l}\mathrm{Tr}\big((Y-JKW)^T(Y-JKW)\big)+\gamma_A\mathrm{Tr}(W^TKW)+\frac{\gamma_I}{(u+l)^2}\mathrm{Tr}(W^TKLKW) \tag{2}$$

where u is the number of unlabeled samples, K is a $(l+u)\times(l+u)$ Gram matrix over labeled and unlabeled data, $Y\in R^{(l+u)\times c}$ represents the corresponding one-hot label with the first l rows corresponding to the labeled data, and the following u rows corresponding to the unlabeled data, which are set to all zero. J is an $(l+u)\times(l+u)$ diagonal matrix given by $J=\mathrm{diag}(1,\ldots,1,0,\ldots,0)$ with the first l diagonal entries as 1 and the rest as 0. L is the Laplacian matrix of the graph constructed from labeled and unlabeled data, and $\mathrm{Tr}(\cdot)$ is the trace operator.

Task 1 (50pt):

Remarkably, the solution for the optimization problem in Equation 2 can be analytically calculated:

$$W^*=(JK+\gamma_A lI+\frac{\gamma_I l}{(u+l)^2}LK)^{-1}Y. \tag{3}$$

Your task:

1. Derive this analytical solution.
2. Implement the derived solution in the `fit` function for the `LapRLS` class (highlighted in TODO) in `semi_sup.py`.

Hint:

1. Start by computing the gradient of the objective function in Equation 2 with respect to W and setting it to zero.

2. For reference, the derivatives of trace are provided below:

$$
\begin{aligned}
&\frac{\partial}{\partial X}\mathrm{Tr}(X^T X) = 2X \\
&\frac{\partial}{\partial X}\mathrm{Tr}(X^T C X) = CX + C^T X \\
&\frac{\partial}{\partial X}\mathrm{Tr}((Y - CX)^T (Y - CX)) = -2C^T (Y - CX)
\end{aligned}
$$

3. The Gram matrix K is a positive definite matrix.

4. $JY = Y, J^T = J, K^T = K$

Task 2 (30pt):

After implementing the LapRLS, you will need to test its performance on the digits and usps datasets. The code for hyperparameter selection has already been provided.

Your task:

1. For each dataset, run LapRLS and RLS 5 independent times.

2. Report the results of each method in each dataset in a table with format mean $\pm$ std and briefly summarize what you observe.

Task 3 (20pt):

In this task, you will use the dimensionality reduction techniques you have learned in class to visualize the data and the prediction results of the two methods.

Your task:

1. Find the implementation of the following dimensionality reduction techniques in scikit-learn: LDA, MDS, Isomap, LLE, and t-SNE (marked with TODO).

2. Run the provided visualization code for the digits dataset and report the visualization results in your homework submission. You will be able to see the misclassified samples (marked as $\times$) in the visualization for RLS and LapRLS.

By completing these tasks, you will gain a deeper understanding of LapRLS and how it compares to RLS. Additionally, you will gain experience with various dimensionality reduction techniques and their visualization capabilities.