

Homework 5

*Hand Out: May.25**Due: June.3*

- Feel free to talk to other students in the class when doing the homework. You should, however, write down your solution yourself. You also must indicate on each homework with whom you collaborated and cite any other sources you use including Internet sites.
- You will write your solution in LaTeX and submit the **pdf file through Gradescope**. You also need to submit the **zipped LaTeX files to CCLE**. We will grade your homework based on the final version of the pdf file submitted to Gradescope. We will not grade the zipped Latex files on CCLE. However, failure to submitting your LaTeX files to CCLE will incur 2 points penalty out of 100 points.
- The homework (both pdf and zipped Latex source files) is due at **1:59 PM before the class**.

The following questions are from *Understanding Machine Learning: From Theory to Algorithms* by Shai Shalev-Shwartz and Shai Ben-David. It can be found here <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/> by courtesy of the authors.

1. (10 points) Exercise **15.1**
2. (10 Points) Exercise **15.2**
3. (10 Points) Exercise **16.2**
4. (10 Points) Exercise **16.3**
5. (10 Points) Exercise **16.4**
6. (10 Points) Exercise **16.6**

In the following two problems, you will implement linear and kernel support vector machine using stochastic gradient descent. Please report your answer of the following questions to Gradescope, and submit your source code to CCLE. Your answer will NOT be graded if we do not see your source code submission on CCLE. Note that, you are allowed to use other programming languages. If so, you need to create an csv data loader and read the data from `./data/*.csv`.

7. (15 Points) Implement the SGD algorithm for solving the optimization problem of Soft-SVM with the objective function

$$J(\mathbf{w}) = \frac{\lambda}{2} \|\mathbf{w}\|^2 + \frac{1}{m} \sum_{i=1}^m \max\{0, 1 - y\langle \mathbf{w}, \mathbf{x}_i \rangle\} \quad (1)$$

- Run the skeleton code, report the training and testing error obtained by the Linear SVM model implemented in scikit-learn.

- Implement SGD algorithm for solving Soft-SVM. Replace line 119-121 of the skeleton code with your implementation.

The objective function (Eq. (1)) has been implemented in the function `LinearSVM.objective`. In this problem, we will keep a running average of updated weights, namely

$$\bar{\mathbf{w}}^{(t)} \doteq \frac{1}{t} \sum_{i=1}^t \mathbf{w}^{(i)}$$

Plot the objective value $J(\bar{\mathbf{w}}^{(t)})$ with respect to the number of iterations t during training. Note that, optimal hyperparameter λ (in Eq. 1) and T (SGD maximum number of iterations) may vary with the implementation. Thus, you may need to tune the hyperparameter to reach good performance.

- Report the training error and testing error on the dataset with the hyperparameter λ and T you used.

8. (15 Points) Implement the SGD algorithm for solving the Soft-SVM optimization problem with RBF kernel. In this problem, we focus on the objective function of the dual Soft-SVM problem,

$$\Theta(\boldsymbol{\alpha}) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_j, \mathbf{x}_i) \quad (2)$$

where the RBF kernel is defined as

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2)$$

- Run the skeleton code, report the training and testing error obtained by the Soft-SVM model with RBF kernel implemented in scikit-learn.
- Implement RBF kernel. That is, given two sets of samples $\mathbf{X} \in \mathbb{R}^{m_1 \times d}$ and $\mathbf{Y} \in \mathbb{R}^{m_2 \times d}$, where d number of dimensions and m_1, m_2 number of samples of \mathbf{X} and \mathbf{Y} respectively, the function `rbf_kernel` returns a matrix $\mathbf{K} \in \mathbb{R}^{m_1 \times m_2}$ with each element $K_{i,j} = \exp(-\gamma \|\mathbf{X}_i - \mathbf{Y}_j\|_2^2)$. Replace line 49-51 in the skeleton code with your implementation.
- Implement SGD algorithm for solving Soft-SVM. Replace line 230-232 of the skeleton code with your implementation.

The objective function of dual SVM (Eq. (2)) has been implemented in the function `RBFSVM.dual_objective`. In this problem, we will keep a running average of the updated alphas,

$$\bar{\boldsymbol{\alpha}}^{(t)} \doteq \frac{1}{t} \sum_{i=1}^t \boldsymbol{\alpha}^{(i)}$$

Plot the objective function of dual SVM $\Theta(\bar{\boldsymbol{\alpha}}^{(t)})$ with respect to the number of iterations t during training. Note that, optimal hyperparameter γ (for RBF kernel), T (SGD maximum number of iterations) and λ (in Eq. 1) may vary with the implementation. Thus, you may need to tune the hyperparameter to reach good performance.

- Report the training and testing error on the dataset with the hyperparameter γ , T and λ you used.