Image Classification using Support Vector Machines and Gaussian Radial Basic Kernel Function

Laleh ROSTAMI HOSOORI
MoSIG Data Science
laleh.rostami-hosoori@grenoble-inp.org

1 Introduction

This approach tries to solve a multi-class nonlinear classification problems. To classify the given data set of images, we use a Support Vector Machine (SVM) and a radial basis kernel function. In this report, first we will present a short introduction to SVMs and then the technical details of the implemented model.

2 Support Vector Machines

Support Vector Machines (SVMs) are supervised learning models in machine learning. They can be useful for classification, regression analysis, and outliners detection. An SVM gets a data set of training examples, all marked with classes (labels), builds model based on the set, and then predicts the class(s) of the new unknown examples. It constructs hyperplane in a high-dimensional space, so that the hyperplane has the largest distance (so-called margin) to the nearest training data point of any class. Generally, the larger the margin is the lower the generalization of the classifier is.

2.1 v-Soft Margin Support Vector Classifiers (SVC)

Practically, it is not possible to obtain a separating hyperplan all the time. Due to the high level of noise existing in the training examples, there might be a large overlap within the classes. Soft margin classifier introduces a decision function by minimizing the following objective function:

$$\tau(w,\xi) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i$$

where constant $C > 0$ and slack variables are

$$\xi \geq 0, i = 1, ..., m$$

subject to

$$y_i.((w.x_i) + b) \geq 1 - \xi_i$$

Applying kernel function, we can rewrite the mentioned equation as Lagrange multipliers. It leads to the following convex optimization problem [1]:

$$c(x, y, f(x)) = \tilde{c}(|y - f(x)|_\epsilon)$$

subject to

$$0 \geq \alpha_i \geq C, i = 1, ..., m\sum_{i=1}^{m}\alpha_i y_i = 0$$

3 Kernel function

Since our classification problem is in higher dimension. We leverage a kernel function to compute the dot product in the feature space. The kernel function is Gaussian Radial Basis function, which satisfies the positive definiteness condition:

$$k(x, x') = exp(-\gamma||x - x'||^2)$$

where $\gamma > 0$.

4 Multi-class classificatiers

The SVMs are originally suitable for binary (two-class) classifications; however, there are several ways to extend them for multi-class problems. We use the ”one-against-one” approach. In this approach, the multi-class problem gets decomposed into several binary sub-problems. In the way that, it trains a binary SVM for any pair of classes and obtains a decision function. It means for a k-class problem, there will be $k(k-1)/2$ decision functions. For predicting the class of a test example, we use a voting approach. Applying the approach, we pick the class with the maximum number of votes for the given test example.

To build a better model, we searched for the proper C and γ and ended up with $C = 100$ and $\gamma = 0.1$.

5 Implementation Details

In a nutshell, here are the list of steps, we took:

1. Load training data set into three arrays: training examples, training labels, and labels
2. Initialize SVC
3. Fit SVC: Load it with the training data set and calculate all possible combinations of label pairs
4. Predict phase:
   (a) Initialize the one-against-one SVC trainer with the RBF kernel
   (b) Find all examples of the two current pair of labels in the training data set
   (c) Train the SVC trainer for the current pair of labels (It will return an one-against-one SVC predictor)
   (d) Predict a label for each test example for the current pair of labels using the predictor
5. Go over all predicted labels: For each test example, the label who occurs the most is the final predicted label
6. Write the result into a file

6 References

[1] Smola A., Scholkopf B., A Tutorial on Support Vector Regression, Kluwer Academic Publishers, 2004