

README

CS3244 Homework 3

Thenaesh Elango, Lee Yi Min, Lim Jia Yee

October 30, 2016

Contents

1	Introduction	2
2	Description	2
2.1	Data Sets	2
3	Learning Models Used	2
3.1	Initial Approach: Support Vector Machine (SVM) only	2
3.2	Second Approach: Recursive Feature Elimination with Cross-Validation (RFECV)	4
3.3	Final Approach: Principal Component Analysis (PCA), SVM, and Neural Networks	5
3.4	Other Approaches	6
3.4.1	Decision Trees, and Adaptive Boosting (Adaboost)	6
3.4.2	Bagging	6
4	File Listing	7
5	Statement of Teams' Independent Work	8

1 Introduction

We can recognise faces after seeing them for a few times. We have essentially “learnt” how to recognise these faces, especially from the more distinctive features of these faces. They include the shape of the face, the curvature of the eyes, the slope of the nose, the thickness of the lips, and so on. We aim to extend this “learning” to computer programs.

2 Description

In the field of machine learning, this problem of face recognition, given previously labelled faces, is a classification problem. These previously labelled data will be referred to as the training data.

The learner, essentially a computer program, attempts to create a general formula based on the training data to “recognise” other face images not in the training data.

2.1 Data Sets

There are 966 training samples and 322 test samples. Each sample has 1850 features, where each feature is simply a pixel. The faces belong to a total of 7 people, and thus there are 7 classes.

3 Learning Models Used

3.1 Initial Approach: Support Vector Machine (SVM) only

The initial models used were SVMs. All 966 training samples and their 1850 features were passed into the SVM. This means there was no feature dimensionality reduction and no cross-validation.

There were a total of more than 80 SVM models with various combinations of arguments, namely kernels, costs, gammas (applicable for radial basis function (RBF) and polynomial kernel only), degrees (applicable for polynomial kernel only).

The kernels used were the linear kernel, the polynomial kernel of degrees 2 to 4 inclusive, and the RBF kernel.

The following ranges are of step size power of 10: The range of costs used was 0.01 to 100 inclusive. The range of gammas used was 0.0001 to 1 inclusive.

For the SVM models using the RBF kernel, all of the test samples were classified under the same class, which seemed to be suspicious. On the other hand,

for the SVM models using the linear or the polynomial kernels, the test samples had a diverse classification, and the SVM models generally agree on the classification for each of the test samples for most of the combinations of the arguments.

A possible explanation for the uniform classification under the RBF kernel could be that RBF is intolerant to slack [1] and thus could not separate the data points which were likely to be close together due to the sheer number of features, many of which could be similar as the human face does not deviate drastically.

As for the linear and polynomial kernels, the classifications were agreeable amongst these models due to:

1. Rather low degrees compared to the number of features
2. Arguments for costs, gamma (if applicable) are within a small range

For any disagreements, the mode of the classification was chosen.

The simple SVM models were decent as the first step, but could have been improved in the following:

1. Ensure the images were oriented in the same angle
2. Gather the more significant features (or pixels) for classification to reduce noise, time, and space complexity
3. Need to try out other models and also think about whether the possible natures of the data set could affect the learning of the different models

As such, in the next batch of models, a new approach with the improvements above in mind will be explored.

3.2 Second Approach: Recursive Feature Elimination with Cross-Validation (RFECV)

The recursive feature elimination (RFE) learning model formulates the weights on each of the features via the training data. The number of features will shrink with every iteration of the entire learning procedure, by removing the features with less significant weights after each iteration.

For RFECV, the optimal number of features is acquired via cross-validation, and the prediction will be based on this pruned set of features.

3.3 Final Approach: Principal Component Analysis (PCA), SVM, and Neural Networks

3.4 Other Approaches

3.4.1 Decision Trees, and Adaptive Boosting (Adaboost)

Since the features are simply about the individual pixels in the face images, the training data used in this section’s approach also underwent PCA transformation in order to attain the more significant features out of all the individual pixels.

A reduced number of features is especially important for decision trees (and Adaboost) because it is expected that each individual pixel will not be effective in maximising information gain. Each pixel is not only affected by the face in the image, but also external noise such as lighting, orientation, and shadow. Hence, an individual pixel by itself may not be effective “decision makers”.

However, despite feature dimensionality reduction with PCA, the F1 scores for both decision trees and Adaboost classifiers lose out to the other classifiers (SVM and neural networks). The F1 score for decision trees has an average of 0.39 while that for the Adaboost is 0.41.

Adaboost is an ensemble method which uses several weak learners (i.e. decision stumps in this context) to combine into a stronger learner. Since Adaboost tries to correct the misclassifications, anomalous or noisy data will steer Adaboost away from the true classification as the learning algorithm attempts to classify these data points with deceiving features [2].

It is expected that the data is noisy, as several external factors affect the quality of the images as discussed earlier in this section.

3.4.2 Bagging

It was noticed that 966 training samples may not be enough for the most accurate classification. In this section, bagging (also known as bootstrap aggregating) was attempted on the SVM model in Section 3.3 Final Approach.

However, there was no significant increase in the F1 score for bagging in SVM, compared to no bagging in SVM (still 0.89).

4 File Listing

5 Statement of Teams' Independent Work

Please initial (between the square brackets) one of the following statements.

[X] We, A0124772E, A0121299A and A0136070R, certify that we have followed the CS3244 Machine Learning class guidelines for homework assignments. In particular, we expressly vow that we have followed the Facebook rule in discussing with others (out of our team) in doing the assignment and did not take notes (digital or printed) from the discussions.

[] We, <substitute your matric numbers here>, did not follow the class rules regarding the homework assignment, because of the following reason: <Please fill in> We suggest that we should be graded as follows: <Please fill in>

References

- [1] Martin C., *What is an RBF Kernel?*, 2012, Feb 6.
- [2] Brownlee J., *Boosting and AdaBoost for Machine Learning*, 2016, Apr 25.