A Study of Principal Component Analysis on Classifiers Using Histogram of Gradients Features

Peter Wei
Carnegie Mellon University
pwei@andrew.cmu.edu

***Abstract*— In image classification, the benefits of Principal Component Analysis (PCA) to a dataset is twofold. Firstly, PCA preserves the most important dimensions while removing dimensions that least contribute to the patterns in data. Secondly, the projection of data along salient dimensions helps draw out patterns in the data. Since individual pixels do not hold much information about objects in an image, using pixels as features loses important information; thus, the experimenter explored the benefits of extracting patterns by extracting a Histogram of Gradients (HOG) from each image, and then performing PCA on the histograms. Finally, the features are classified by two different classifiers, multinomial logistic regression (MLR) and support vector machines (SVM) in order to measure objectively the benefits of PCA. The experimenter found that PCA improved the performance of these classifiers in almost all cases, up to 17.7%.**

I. INTRODUCTION

Principal Component Analysis (PCA) is an important method in machine learning due to it's twofold nature. PCA reduces the dimensionality of the dataset, which takes the dimensions that encode the most important information and removes the dimensions that encode the least important information. By reducing the number of dimensions, the data utilizes less space, thus allowing classification on larger datasets in less time. Further, by taking only the salient dimensions, PCA projects the dataset onto dimensions that hold the most meaning, thus drawing out patterns in the dataset.

In the classification experiments, the experimenter wanted to classify images from a subset of the CIFAR-10 image dataset. One of the prominent ideas in computer vision is extraction of features from an image; this is because individual pixels often do not encode as much information as a pixel's interactions with it's neighboring pixels. In order to extract these features, the experimenter first extracted a Histogram of Gradients (HOG) from each image to serve as the initial features. These features are then reduced in dimensionality by PCA to further focus on salient features in each image.

Finally, two classifiers, multinomial logistic regression (MLR) and support vector machines (SVM) are trained on the dimensionality reduced features. The extraction of features allowed each classifier to perform better than without extraction of features; even reducing the dimensionality of the features by 70% still allowed feature extraction to perform better than the baseline without HOG or PCA. In the best case, PCA was found to improve the multinomial logistic regression classifier with HOG feature extraction by 17.7%.

II. BACKGROUND

A. Multinomial Logistic Regression

The experimenter used 2 baselines for testing the objective benefits of PCA. The first is multinomial logistic regression, which utilizes a logistic function to predict the outcome of an observation. The parameters are then put into a linear combination for each test observation to determine the class of that observation.

B. Support Vector Machines

The second baseline used to test the benefits of PCA is support vector machines.

Support vector machines attempt to maximize the separation between the categories; the ideal solution is found by a linear gap with no observations in the gap, that is as wide as possible and that separates two classes in an n-dimensional space. The space is often transformed by a kernel method to allow for non-linear classification. The data is classified by which side of the gap it lies on; multi class problems are usually reduced to multiple binary SVM classifications (SVM classification with only 2 classes).

C. TotalBoost

A side study was also performed to provide supporting data for the benefits of PCA. This study utilized an ensemble method called TotalBoost, which incrementally builds a model where each model emphasizes the misclassifications of the previous model. This experiment focused on TotalBoost, which often requires less iterations than AdaBoost (the more commonly used ensemble method). The use of TotalBoost over AdaBoost drastically decreased the amount of time required for testing, from approximately 5 minutes to 2 minutes without PCA [1].

III. METHODS

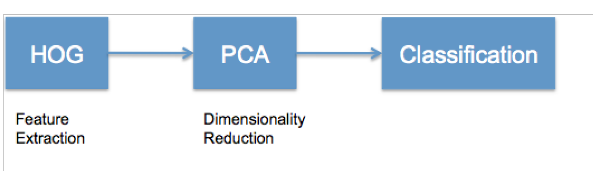

Fig. 1: General pipeline for classification

A. Histogram of Gradients

In order to improve the baseline, the experimenter chose to use methods to extract salient features from the images. Using features instead of pixels for classification could encode more information about the object in the image. Before running multinomial logistic regression, the experimenter ran the MATLAB implementation of extracting a Histogram of Gradients (HOG) from images. This helped to extract features about the object from the image. Because the images are relatively small (32x32x3), finding other types of features such as SURF features is not guaranteed to succeed; in fact, on some of the images in the dataset, MATLAB's find SURF features functions return no SURF features. HOG consistently returns a vector of features, and thus is a good feature type for extraction from small images.

HOG computes a feature descriptors in a grid in the image. The grid consists of points that are uniformly spaced throughout the image, and the gradients end up overlapping for greater accuracy. HOG relies on the idea that an object's appearance can be broken up into smaller images and that local gradients can encode a significant amount of information. Together, the local gradients (especially with the overlapping property) can encode more information about the object than just the pixels, which do not encode any local information except for the pixel's current location.

B. Principal Component Analysis

Once the Histogram of Gradients have been computed for each image, Principal Component Analysis helps to further reduce the dimensionality of the features. This allows for additional focus on the underlying patterns in the images. In addition, PCA reduces the amount of memory used by the data, thus allowing for faster classification and better memory usage. PCA returns principal components, and reconstruction with the first n components recovers the original features now projected onto the new dimensions.

C. Parameters

Changing the regularization parameter, λ, also affects the performance of the classifiers on the test data. Without PCA or HOG, multinomial logistic regression and support vector machines perform better on the test data with a high λ, due to prevention of overfitting to the training data. However, when adding HOG and PCA the classifiers no longer overfit to the training data (as seen by the $\sim$40% training accuracy). Thus, reducing the

λ parameter improves the test classification accuracy of the classifiers up to a certain point.

IV. RESULTS

A. Experiments

For the first experiment, the experimenter classified SVM and MLR with and without PCA.

Num Components	Train Accuracy	Test Accuracy
No PCA	**100%**	42.9%
50	63.1%	41.1%
100	61.1%	41.3%
200	78.3%	42.8%
300	70.0%	41.5%
500	89.7%	**43.1%**

TABLE I: MLR with and without PCA

Num Components	Train Accuracy	Test Accuracy
No PCA	**53.3%**	37.0%
500	49.0%	**37.4%**

TABLE II: SVM with and without PCA

For the second experiment, the experimenter classified MLR and SVM with HOG, with and without PCA.

Num Components	Train Accuracy	Test Accuracy
No PCA	28.73%	28.4%
100	29.8%	29.6%
200	30.3%	30.0%
298	**46.6%**	**46.1%**
300	44.6%	43.1%

TABLE III: MLR and HOG with and without PCA

Num Components	Train Accuracy	Test Accuracy
No PCA	45.7%	43.3%
300	**51.2%**	**46.7%**

TABLE IV: SVM and HOG with and without PCA

The last experiment the experimenter performed was with changing the regularization term λ.

λ	Train Accuracy	Test Accuracy
300	45.73%	43.3%
300 with PCA	51.3%	46.6%
5	57.7%	**48.2%**
5 with PCA	61.8%	46.9%
1	**62.4%**	47.2%
1 with PCA	64.2%	45.3%

TABLE V: SVM and HOG with varying λ

In addition to these experiments, the experimenter ran a timing experiment using TotalBoost, an ensemble classifier. The experimenter ran the experiment with and without PCA to check the accuracy. Once again, the experimenter found a noticeable increase in accuracy with TotalBoost run with PCA than without. However, the accuracy was much lower than the other classifiers, and thus is not considered to be valid data.

B. Discussion

PCA allows the classifier to behave differently when given HOG or original image data. When given the original image data, PCA does not give a significant improvement over the baseline (no PCA). However, when given local feature information, such as from HOG, PCA significantly improves the test accuracy (improvement of 17.7% for MLR and HOG). An improvement of 3.4% classification accuracy for SVM and HOG is also a noticeable difference.

Initially, it seems strange that decreasing the regularization term increases the accuracy. However, since the training accuracy is similar to the test accuracy, one could surmise that the classifiers are not overfitting. Since regularization seeks to introduce information to reduce overfitting, it is possible that a large λ could interfere with classification especially if there is little overfitting.

Nevertheless, it is important to have a certain amount of regularization. When the regularization term gets too low, the training accuracy does continue to improve (as would be the case for overfitting) but the testing accuracy decreases for both the cases with PCA and without.

V. CONCLUSIONS

The two experiments show a contrast when using PCA. The first experiment uses pixels as the features; the second experiment uses HOG as the features. HOG theoretically encodes different local information than pixels, and from the results, PCA behaves differently depending on the encoded data. With HOG, the classifiers improve significantly on the test data with PCA, while without HOG, the classifiers improve slightly on the test data with PCA. In addition, decreasing the λ also increases the accuracy up to a certain point.

The results of the experiments show that PCA is an overall effective method for improving classification accuracy. It seems that pairing PCA with an algorithm for extracting salient information about the dataset, such as in [2], performs better than PCA alone. In the experiments presented, pairing PCA with HOG increased the performance of both SVM and MLR. PCA is an important algorithm in machine learning, and the results shown in this paper support the idea that PCA can improve the test accuracy for a variety of systems.

VI. BIBLIOGRAPHY

[1] Lemaître, Guillaume, and Miroslav Radojevic. ”Directed Reading: Boosting algorithms.” (2009).

[2] Jiang, Jiafu, and Hui Xiong. ”Fast pedestrian detection based on hog-pca and gentle adaboost.” Computer Science & Service System (CSSS), 2012 International Conference on. IEEE, 2012.

VII. APPENDIX

All code, experiments, data collection, and paper writing was done by Peter Wei (pwei).

In this project, I learned a potential strategy for maximizing the effectiveness of PCA. In addition, I learned about overfitting in practice and the importance of setting a good regularization term. I learned about the pros and cons of a multitude of machine learning algorithms, and some of the situations which suit them best. I learned the importance of analyzing data for patterns and the importance of tuning parameters to find the best set for testing.

If I were to do this project again, I would do it with a partner. A two person team could collect more data and analyze more classifiers, and possibly run the original dataset through a different feature extraction algorithm (such as SIFT).