A comparative study for using PCA, LDA, GDA, and Lasso for dimensionality reduction before classification algorithms

Duyemo Anceaux -4919327
duyemoa@gmail.com

Responsible professor: Asterios Katsifodimos
Supervisor: Andra Ionescu

1. Background

- Curse of dimensionality: to many features to process
- Dimensionality reduction can be used to reduce number of features
- Previous studies have shown no one algorithm in universally the best at dimensionality reduction[3][4]

2. Research question

What are the effects of PCA, LDA, GDA, and Lasso for dimensionality reduction on classification algorithms?

3. Preliminaries

- PCA: find directions of maximal variance[2]
- LDA: minimize within-class scatter while maximzing between-class scatter[2]
- GDA,: LDA with different matrices per class and kernel function[2]
- Lasso: Linear model with a penalty function with weight α [1]

4. Metrics

- Transformation time
- #Features retained
- Accuracy

5. Results

PCA	LDA	GDA	Lasso
00:00.89	00:01.38	49:13.65	00:01.15

Table 1: The time it took each of the algorithms the transforms a dataset into a new dataset with less dimensionality

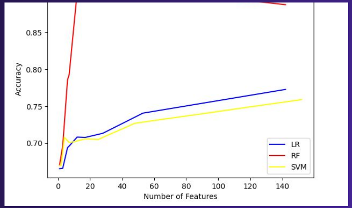

Figure 1: Effect of Number of features left by Lasso on the accuracy of a dataset with a lot of training data

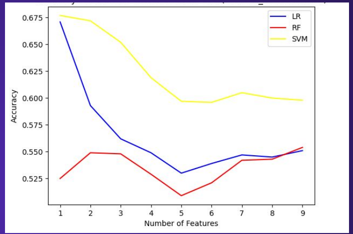

Figure 2: Effect of Number of features left by GDA on the accuracy of a dataset with less training data

	LR			RF			SVM		
	acc<10	best acc	best #features	acc<10	best acc	best #features	acc<10	best acc	best #features
PCA	0.667	0.736	110	0.798	0.809	19	0.670	0.732	110
LDA	0.759	0.759	9	0.798	0.798	9	0.750	0.750	8
GDA	0.745	0.745	7	0.762	0.762	7	0.752	0.752	8
Lasso	0.696	0.773	142	0.793	0.909	53	0.702	0.759	142

Table 2: Statistic about the combinations of classification and dimensionality reduction methods on dataset with lots of training data

	LR			RF			SVM		
	acc<10	best acc	best #features	acc<10	best acc	best #features	acc<10	best acc	best #features
PCA	0.676	0.678	15	0.684	0.685	18	0.676	0.689	39
LDA	0.671	0.671	1	0.550	0.550	2	0.637	0.637	1
GDA	0.671	0.671	1	0.549	0.549	2	0.677	0.677	1
Lasso	0.711	0.712	19	0.698	0.742	19	0.707	0.707	10

Table 3: Statistic about the combinations of classification and dimensionality reduction methods on dataset with less training data

6. Conclusion

-Changes in number of features and classes does not change relative performance of algorithms
-Number of features does change relative performance of algorithms
-LDA in better then GDA (using linear regression)
-LDA and GDA are better when using LR or SVM and training set is sufficiently large
-Lasso and PCA are better when using RF or if the training set becomes small

Refrences

[1]V. Fonti and E. Belitser, "Feature selection using lasso," VU Amsterdam research paper in business analytics, vol. 30, pp. 1–25, 2017
[2] P. Yu, P. Yu, and D. Xu, "Comparison of pca, lda and gda for palmprint verification," in 2010 International Conference on Information, Networking and Automation (ICINA), vol. 1, 2010, pp. V1–148–V1–152.
[3]A. Babjac, T. Royalty, A. D. Steen, and S. J. Emrich, "A comparison of dimensionality reduction methods for large biological data." New York, NY, USA: Association for Computing Machinery, 2022. [Online].
[4]A. Martinez and A. Kak, "Pca versus lda," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228–233, 2001.