SA Youth Labour Market Participation: Applying sPCA as a pre step before modelling.

Bogatsu M. Siphiwe

March 22, 2024

Introduction

Youth people in South Africa continue to be unable to find work and withdrawing from the labour force entirely, with youth unemployment and underemployment overwhelmingly standing at a rate above that of the national average. According to StatSA (2024), for the first quarter of 2022 the unemployment rate was 63,9% for those aged 15-24 and 42,1% for those aged 25-34 years, while the current official national rate stands at 34,%. Being able to build statistical models of reality that can predict, and understand, which young people will find employment and which ones will require additional help, helps promote evidence-based decision-making, supports economic empowerment, and allows young people to thrive in their chosen careers. However, youth survey datasets present - useful in building these models - are high dimensional in their nature. This presents a critical challenge of how to reduce the dimension of these datasets and the ability to extract relevant features as means towards building robust models to predict which young people will find employment.

As classical tool, principal component analysis (PCA) has been widely used approach for data dimensionality reduction, and data processing as a pre-step in fitting machine learning models (Zou et al, 2006). From a multivariate data set, it finds a few uncorrelated linear combinations of the original data, called principal components (PC), with an intention of variance maximization in the new, low dimensional space. In essence, each PC is a weighted average of all predictors with a weight (loading coefficient) attached to each predictor. This approach has led to fruitful stories in numerous applications in biology, engineering, and social science.

Often times, PCA is computed through the singular value decomposition (SVD) of the original data matrix. For instance, let the data X be a n x p matrix, such that n and p are number of observations and the number of variables, respectively. Assuming the data matrix is centered, the SVD of X is:

$$X = UDV^T$$

From there, PC are computed as $Z = UD$, the loading coefficients of the PC, as the basis of the low dimensional space, are represented by the columns of V, and the sample variance of each PC are in the diagonal matrix D. The 1^{st} PC explains the highest total variation, the 2^{nd} PC explains the 2^{nd} largest variation, and so on. The prestige of PCA lies in two its properties:

1. Its ability to use few PCs to sequentially capture maximum variability among the columns of the original data X, thus ensuring minimal loss of information.

2. PCs are uncorrelated with one another.

Therefore, this puts us in a position to easily apply standard statistical methods, such as logistic regression or linear regression models, using a few, say 4 PC from a highly dimensional data.

However, it has an obvious drawback. That is, PCA reduces data into lower dimensional space into PCs, to maintain the same structure as the original data. Each PC uses *all* the variables regardless if a predictor is a noise - meaning no true effect in the PC; most of the variable loadings are non-zero (Zou et al, 2006). This raises serious concerns, because would not it better to exclude the predictor during the construction of PCs such that they become more robust (That is, concentrating the variability into a few PCs that easy to interpret) ? The result of using all predictors, including the noisy ones, will generate PCs likely contaminated with noise such that the resulting PCs may deviate significantly from the original data (Hsu, Huang & Chen, 2014). This problem becomes exacerbated when the number of predictors is sufficiently large, thus making it difficult to interpret the derived PCs. Moreover, in some applications, having non-zero loadings can cause a direct cost (for instance, transaction costs in finance) (d'Aspremont et al, 2008).

Therefore, the aim of this study is to derive sparse principal components - a set of sparse vectors that explain maximum possible amount of variance in the SA youth data to address this issue. Sparse PCA merges the benefits of traditional PCA, namely data reduction, with sparsity modeling, which removes ineffective variables from the PCA model by reducing their loadings to zero (Hsu, Huang & Chen, 2014).

This paper is organised as follows. In Section 1, I begin by formulating the sparse PCA problem through three approaches: variance maximisation, reconstruction error and singular value decomposition. I describe in depth the SA youth data, including exploration and data pre-processing in Section 2. Finally, SA youth data is used to illustrate the potential application of sparse PCA.

1 Methods

Variance maximization approach (VM)

This approach reduces the data into a low dimensional space with an aim of preserving variation of original sample points as much as possible. As highlighted above, the first loading coefficient vector is V_1, and the corresponding PC is XV_1. The maximization of the variance of XV_1 can be expressed as the form:

$$\max_{V_1}(V_1'X'XV_1) \tag{1}$$

$$\text{subject to } V_1'V_1 = 1 \tag{2}$$

It becomes a sparse PCA after imposing a L1-norm penalty function:

$$\max_{V_1}(V_1'X'XV_1) + \lambda_1\|V_1\|_1 \tag{3}$$

$$\text{subject to } V_1'V_1 = 1 \tag{4}$$

where $\|V_1\|_1 = \sum_{i=1}^{p} |V_{i1}|$. The λ_j, $j = 1, 2, \cdots, k$, is a penalty parameter that controls the shrinkage amount on each PC. The higher the value of λ_j, the greater the amount of shrinkage to zero. To implement an VM algorithm an R package, *pcaPP*, is available (Croux et al, 2014)

Reconstruction error minimization (REM)

Zou, Hastie & Tibshirani (2006) reconstruct the product of the loading coefficient matrix, $V'V$, into two matrices (A, B) - both being $p \times k$ matrices - and then add an L2-norm penalty on B, such that the PCA model becomes:

$$\min_{A,B} \sum_{i=1}^{n} ||x_i - AB'x_i||^2 + \lambda \sum_{j=1}^{k} ||B_j||^2 \tag{5}$$

$$\text{subject to } A'A = I_k \tag{6}$$

where $||B_j||^2 = \sum_{t=1}^{p} (\sqrt{B_{tj}})^2$ and λ is the penalty parameter. The j^{th} loading is $V_j = \frac{B_j}{||B_j||}$, $j = 1, 2, \cdots, k$.

An L1-norm penalty on B is added to obtain sparse loadings to impose sparseness in the model, such that the model becomes:

$$\min_{A,B} \sum_{i=1}^{n} ||x_i - AB'x_i||^2 + \lambda \sum_{j=1}^{k} ||B_j||^2 + \sum_{j=1}^{k} \lambda_j ||B_j||_1 \tag{7}$$

$$\text{subject to } A'A = I_k \tag{8}$$

Note that there is a common λ used for all PCs, but different λ_j's are used for penalizing the loadings of different PCs. A R package, *elasticnet*, is available to perform the REM (Zou, Hastie & Tibshirani, 2006).

Singular Value Decomposition (SVD)

This approach involves the estimation of U, D, and V in the following formulation:

$$\min_{U,D,V} ||X - UDV'||^2 \text{ subject to } U'U = I_k \text{ and } V'V = I_k \tag{9}$$

To promote sparse loadings, an L1-norm penalty in equation (9), such that the penalized formula becomes:

$$\min_{U,D,V} ||X - UDV'||^2 + \sum_{j=1}^{k} \lambda_j ||V_j||_1 \text{ subject to } U'U = I_k \text{ and } V'V = I_k \tag{10}$$

where λ_j is the penalty parameter for each component.*PMD* is available as an R package to implement the SVD approach to sparse PCA.

Note that the choice of a particular value of λ_j is done through the Tradeoff Product Optimization (TPO) criterion. TPO maximises the area under the tradeoff curve shown in Figure 5 - in particular it maximises the explained variance multiplied by the number of zero loadings of a particular component. In the context of sparse PCA, it refers to the loss of explained variance vs. the improvement of sparseness. So a range of different values for λ_j are tested, and then the best choice of λ_j for each PC is chosen as a result of maximising the area under the tradeoff curve. These λ_j are used for both VM and REM approaches considered in this study.

In this section, I have discussed the methods to be used to derive the sparse principal components, and how I got to determine the choice of sparseness parameters. Next, I detail the data source and pre-processing techniques before applying the methods in R.

Data Pre-processing and Exploration

This section details the data source used to conduct the study, the pre-processing and feature engineering techniques deployed before the application of PCA and sparse PCA.

Data Used

The data was collected by a company situated in Stellenbosch, Predictive Insights, through 4 rounds of a survey of youth in the South African labour market, conducted at 6-month intervals. Each youth in the dataset was surveyed one year prior to the follow-up survey, which specifically asked whether the youth is currently employed. Table 5 in the appendix of this report, shows the description of all the variables considered in this study.

In the following subsection I show the traits of the target variable (whether a youth individual is employed) and some of the interesting predictors.

Exploratory Data Analysis

Target variable breakdown

Figure 1 displays the breakdown of the youth individual employment status (0: unemployed and 1: employed) after a year since the baseline survey. It can be seen that 2927 youth individuals out of 4020 survey were still unemployed a year later. This means that only 27.2% of the youth individuals got employed within the SA labour market. This is not a surprising observation as the labour market in SA continues to offer less opportunities for the youth to derive livelihoods.

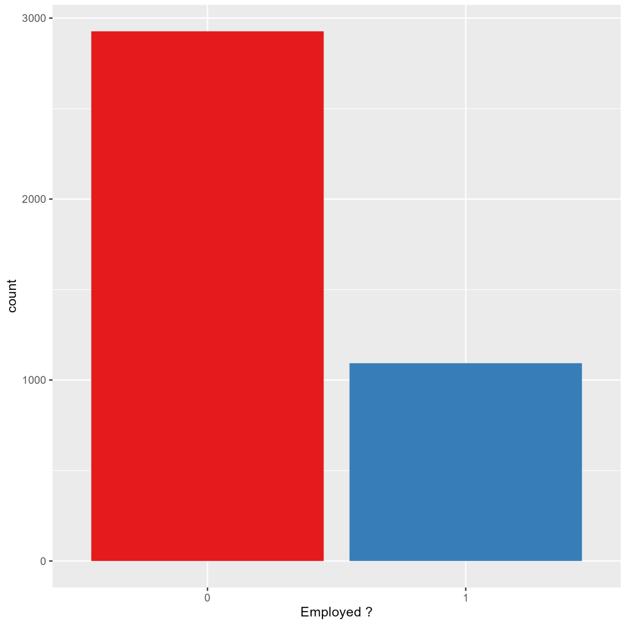


Figure 1: SA youth labour market data: Illustration of the youth individual employment status. The target variable.

Age of the youth individuals

Figure 2 shows the distribution of age for both the unemployed and employed individuals after a year since the baseline survey. It is clear that these groups follow the same distribution shape - skewed to the right; the individuals in their early 20s have the highest participation in the SA labour market, and then it gradually subsides with the years.

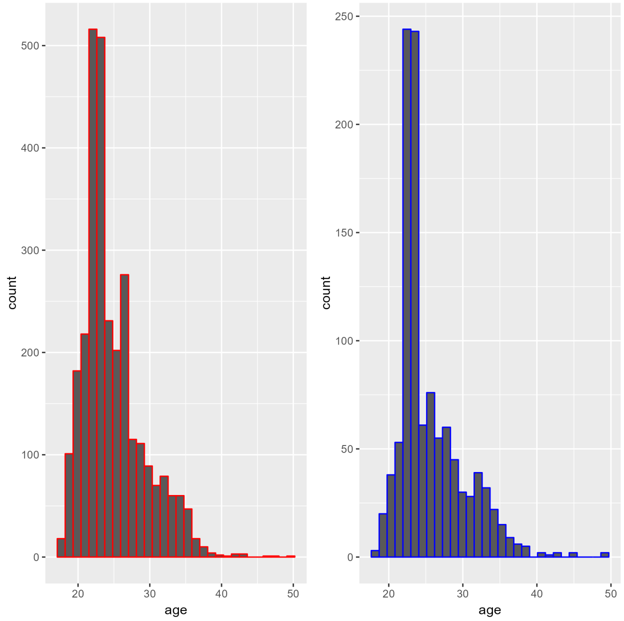


Figure 2: SA Youth Data: The distribution of the youth individuals' age. Red for unemployed individuals and blue for employed individuals, after a year since the baseline survey.

Education Attainment

Figure 3 shows the education attainments of the youth individuals in the labour market. It is evident that youth individuals seeking work are quite substantial across all education attainments relative to the employed ones. Matric attainment is higher than diploma and degree attainments in the SA labour force - Over 2000 youth individuals have attained a matric certificate have remained seeking for after a year since the baseline survey. While on the other end of the spectrum, there is only 65 and 71 youth individuals who found work shortly after a year with a degree and diploma, respectively.

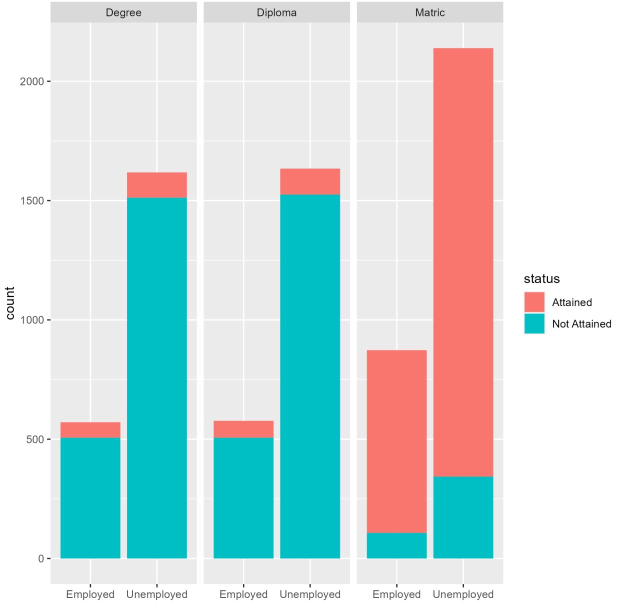


Figure 3: SA Youth Data: Illustration of education attainment of the youth individuals with their employment outcomes

Geography

Figure 4 shows the geographical distribution of the youth individuals in the study alongside their employment outcomes after a year of the baseline survey. Clearly, the study focused more on urban dwellers than both suburban and rural dwellers. In the urban areas, for every 1 employed youth individual there is 2.44 unemployed ones; rural is 1 : 3.27, and suburb 1 : 3.61. That means living in the urban areas of SA, one is more likely going to find work in a year, than anywhere else considered.

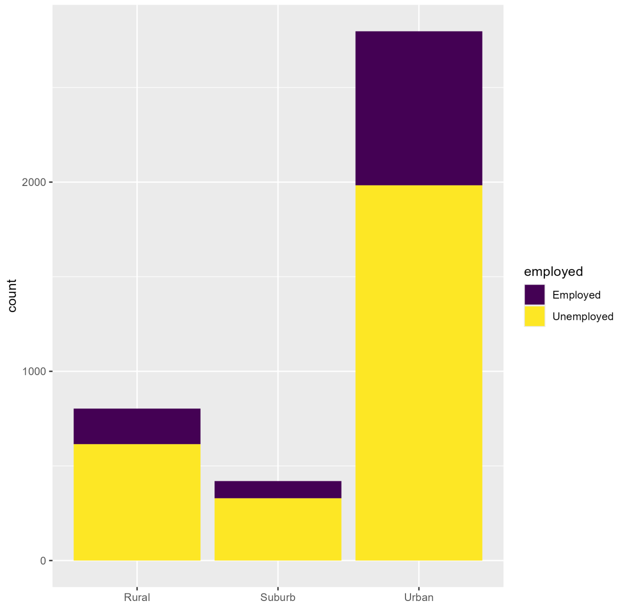


Figure 4: SA Youth Data: Illustration of geographical location of the youth individuals with their employment outcomes

The source of the data and data itself of this project have been described.

Pre-processing

This section details how the predictors were scaled and encoded and how missing values were handled in this project.

Scaling and encoding

Numeric predictors were standardized just before applying sparse PCA. Standardising numeric features involves transforming the values of each variable so that the values of variable have a mean is 0 and the standard deviation is 1.

Notably, *Science*, *Home lang*, *Additional lang*, *Maths* and *Mathlit* subject were initially recorded as intervals between $0\% - 100\%$. Thus, a median of an interval was used as the subject result for each youth individual. Then, numeric normalization was applied.

The categorical variables were encoded using the weight of evidence encoding (WOE). This is a common approach for handling categorical variables within the credit risk and financial industries (Stone, 2020). But, I thought this technique could be handy in our study area. Essentially, WoE encoding scales the levels of a categorical predictor variables based on their relationship with the target variable. In our case, it encodes against the whether or not the young person is employed after a year of survey.

Also, WOE encoding places missing values into a category and assigns a scaled value to them; making it useful in handling missing values in categorical data. To compute the WOE for this study:

$$WoE = ln(\frac{POE}{POU}) \tag{11}$$

where POE and POU represent the probability of the youth individual being employed and probability of the youth being unemployed, respectively. As an example, the WOE scores of the categorical feature, *female*, is shown in the Table 1.

Variable	WOE score
Male	25.14
Female	-21.55

Table 1: Weight of Evidence (WOE) score for the *female* variable.

Missing values

To deal with missing values, I firstly decided to exclude 1008 youth individuals who have missing values for all the education attainment features(*Matric*, *Diploma* and *Degree*), simultaneously. I assumed that the missing values are missing not at random, thus imputation technique of replacing the missing values with the mean of the continuous feature and the mode for categorical variables was not going to be successful.

I explored the k-nearest neighbours (KNN) method to impute the remaining missing values. The KNN approach involves filling the missing values of a particular row with the average value of the equivalent variable from the row's K "most similar" neighbours (Stone, 2020). Note that before the KNN model was developed to replace the missing values, each feature was normalised and scaled. Here is the KNN algorithm I followed:

- Arbitrarily select the number of nearest neighbours to be considered for each youth individual to 4.
- Check if the youth individual row had any missing values. If the row did not have missing values move to the next row, if it did then continue to the steps below.
- Calculate the Euclidean distance between the youth individual row under consideration and every other youth individual row.
- Identify the 4 closest youth individuals based upon Euclidean distance
- Fill each missing value with the mean value of the respective variable taken from the youth individual's 4 nearest neighbours

In summary, this chapter details the data source for this study, conducts a short exploratory data analysis on the target variable and a few interesting predictors, and then details the pre-processing techniques used prior to modelling, which includes scaling and encoding and handling missing values.

Results

For this purpose, I only included only the first ten principal components. I set

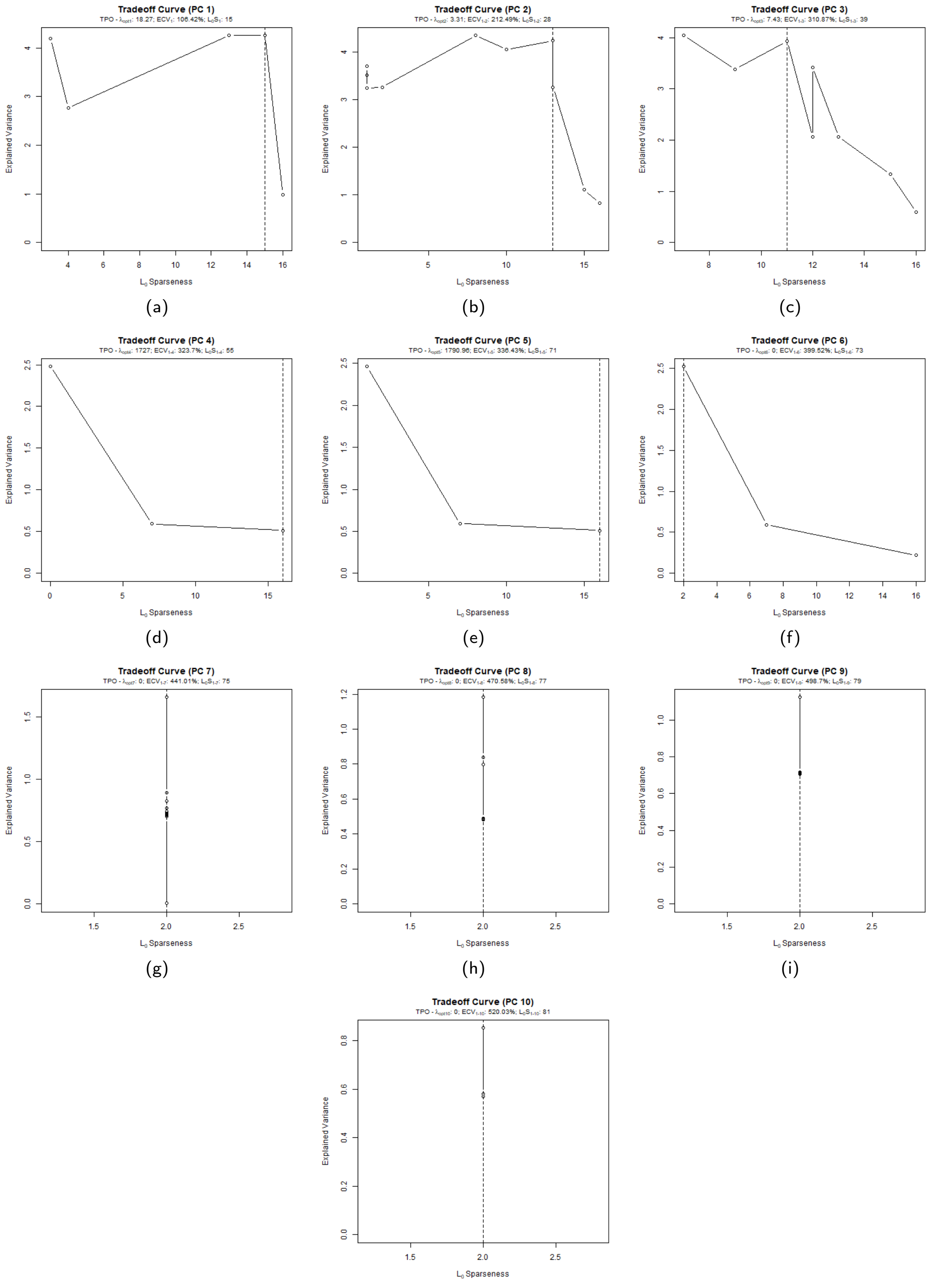
$$\lambda_j = (18.27, 3.31, 7.43, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00)$$

and $\lambda = 0$, chosen according to Figure 5, such that the sparse loadings explain the maximum possible variability as much as an ordinary PC did under PCA. Table 2 shows the loadings from the PCA constructed using the SVD approach. Table 3 and Table 4 show the obtained sparse loadings and the corresponding variances constructed using the VM and REM approach respectively. Since it never compromises for sparseness, PCA will have largest explained variance captured by the PCs compared to the sparse PCA approaches. Compared with the REM approach, the VM approach accounts for relatively small explained variance (72% vs 78%). In fact, the REM approach is not far off from the PCA (80%) in terms of variability explained.

To demonstrate how sparse PCAs reduce the loading coefficients towards zero, I contrast the loading coefficients of each sparse PCA with those of the traditional PCA in Figure 6. The findings suggest that loading coefficients approaching zero in the traditional PCA are likely to be further reduced to zero by sparse PCAs. This is shown by REM (black) in the 7th and 9th PCs.

Notice that while *all* loadings are non-zero in PCA in Table 2, both the sparse PCA approaches shrink a significant number of loadings; REM shrinks at total of 40 loadings to zero in the entire structure; and the VM achieves a highest number of loadings shrunk to zero at 99. Thus, this leads to a decrease in the total explained variability for both sparse PCA approaches - the decrease is small and relatively benign for the REM (2.9%) than it is for VM (8%), as expected.

A subset of predictors are considered as *important* only if they are from a (sparse) principal component that could explain a large part of the total variance of the SA youth data (Zou, Hastie & Tibrashani, 2006). In our case, about 70% of these 17 predictors are sufficiently able to construct the first (sparse) principal component with no loss in explained variance (14.6%), under the REM approach. VM takes an affordable loss of explained variance (14% to 10%) to use slightly half (53%) of predictors to construct the first (sparse) principal component. *Matric*, *Languages*, *Maths*, *Tenure*, *Province*, *Schoolquintile* are observed to be the important variables in the construction of the first few (sparse) principal components that atleast explain 50% variability in the structure of the SA youth data.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j)

Figure 5: SA youth data: The curves show the percentage of explained variance (PEV) as a function of λ_j. The vertical broken lines indicate the choice of λ_j used in both SPCA analysis approaches.

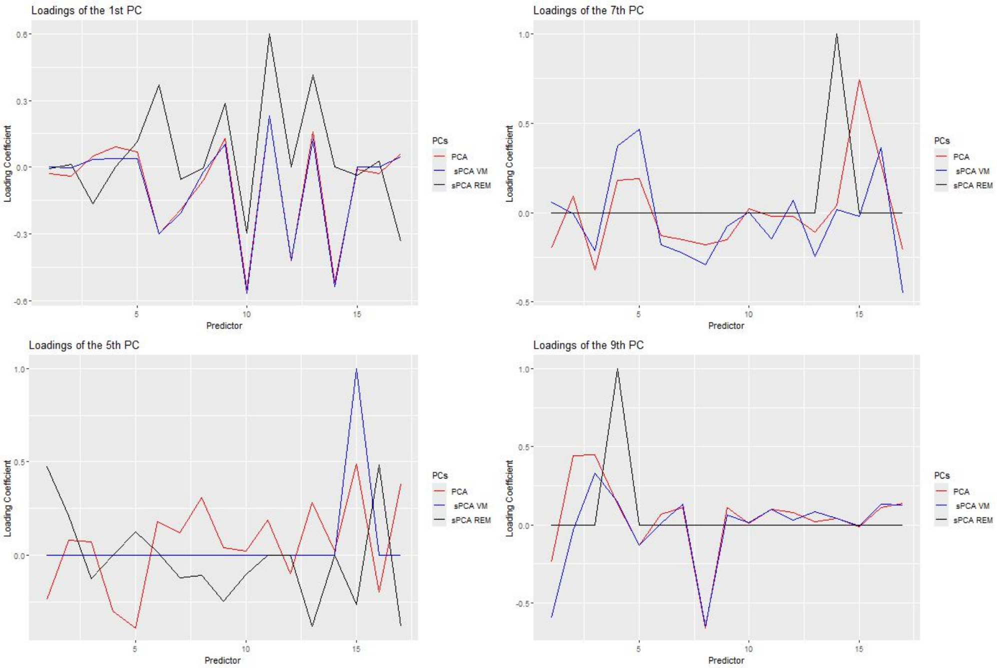


Figure 6: SA youth data: Illustration of individual loadings in selected PC (1,5,7,9) for standard PCA, VM and REM

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Round	-0.03	0.08	0.14	-0.49	-0.24	0.33	-0.20	-0.39	-0.24	-0.12
Status	-0.04	0.09	0.18	-0.53	0.08	0.23	0.09	-0.19	0.44	0.31
Tenure	0.05	0.11	-0.17	-0.09	0.07	-0.09	-0.32	0.22	0.45	0.33
Geography	0.09	0.06	0.30	-0.14	-0.30	-0.51	0.18	0.16	0.13	-0.05
Province	0.07	0.11	0.29	-0.23	-0.39	-0.31	0.19	0.19	-0.13	0.05
Matric	-0.30	-0.45	0.32	-0.06	0.18	-0.04	-0.13	0.09	0.07	0.04
Degree	-0.19	0.10	0.16	-0.10	0.12	-0.04	-0.15	-0.06	0.11	-0.76
Diploma	-0.06	0.02	0.08	-0.30	0.31	-0.04	-0.18	0.34	-0.66	0.25
School quintile	0.13	0.12	0.42	0.13	0.04	0.05	-0.15	-0.04	0.11	0.07
Math	-0.55	0.25	0.13	0.13	0.02	-0.02	0.02	0.02	0.01	0.06
Math lit	0.23	-0.56	0.19	-0.07	0.19	-0.05	-0.02	0.03	0.10	-0.09
Additional_language	-0.42	-0.46	-0.14	-0.14	-0.10	-0.14	-0.02	0.02	0.08	0.01
Home_language	0.16	0.13	0.56	0.29	0.28	0.13	-0.11	0.05	0.02	0.04
Science	-0.52	0.21	0.11	0.14	0.02	-0.03	0.04	0.01	0.04	0.13
Female	-0.01	0.05	-0.01	-0.14	0.49	-0.05	0.74	-0.15	-0.01	-0.01
Birth Month	-0.03	-0.03	0.03	-0.05	-0.20	0.60	0.27	0.68	0.11	-0.18
Age	0.06	0.29	-0.21	-0.32	0.38	-0.24	-0.21	0.29	0.14	-0.25
Variance (%)	14.67	11.10	10.67	7.76	7.07	6.61	6.13	5.64	5.49	4.96
Cumulative Variance (%)	14.67	25.78	36.44	44.21	51.28	57.90	64.03	69.67	75.16	80.12

Table 2: SA youth data: Loadings of the first 10 Principal Components by PCA

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Round		0.26	0.38		0.48	-0.21				-0.22
Status		0.27	0.44	0.27	0.21	0.13				
Tenure	-0.16	-0.14	0.20		-0.13	-0.26				0.33
Geography									1.00	
Province	0.11	0.16	0.36	-0.28	0.13	0.13				-0.36
Matric	0.37	0.35	-0.32	0.37		-0.24				
Degree		0.37			-0.12	-0.13				
Diploma		0.18	0.15	0.37	-0.11	-0.27				
School quintile	0.29	0.19	0.26	-0.37	-0.25					0.15
Math	-0.29	0.59	-0.34		-0.11					-
Math lit	0.60	-0.26		0.32		-0.11				
Additional_language								1.00		
Home_language	0.42	0.25	0.13	-0.27	-0.38					0.22
Science							1.00			
Female				0.45	-0.26	0.73				-0.19
Birth Month					0.48	0.34				0.72
Age	-0.34		0.40	0.23	-0.38	-0.16				0.26
Variance (%)	10.50	9.70	7.76	7.2	6.89	6.06	6.06	6.06	6.06	5.79
Cumulative Variance (%)	10.50	20.19	27.96	35.15	42.04	48.10	54.16	60.22	66.27	72.06

Table 3: Table 2. SA Youth Data: Loadings of the First Ten Modified PCs by VM approach. Empty cells have zero loadings.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Round		0.04	0.03			0.07	0.06	-0.12	-0.60	-0.68
Status			0.02	-1.00				0.03	-0.03	-0.07
Tenure	0.03	0.10	-0.15			-0.13	-0.21	0.22	0.33	-0.15
Geography	0.04	0.05	0.27			-0.54	0.37	0.15	0.15	
Province	0.04	0.10	0.24			-0.39	0.47	0.21	-0.13	-0.05
Matric	-0.30	-0.47	0.31			-0.03	-0.18	0.13		-0.04
Degree	-0.21	0.07	0.15			-0.07	-0.23		0.13	-0.51
Diploma			0.06			-0.11	-0.29	0.45	-0.65	0.39
School quintile	0.10	0.12	0.44			0.05	-0.08	-0.07	0.06	-0.05
Math	-0.57	0.24	0.12			0.05		-0.02	0.02	0.07
Math Lit	0.23	-0.57	0.21			-0.06	-0.15	0.06	0.10	-0.08
Additional_language	-0.41	-0.48	-0.17			-0.15	0.07	0.05	0.03	-0.05
Home_language	0.13	0.14	0.63			0.22	-0.25		0.08	0.08
Science	-0.54	0.20	0.11			0.04	0.02	-0.03	0.04	0.10
Female					1.00					
Birth Month		-0.02				0.59	0.36	0.68	0.13	-0.15
Age	0.05	0.26	-0.19			-0.31	-0.45	0.42	0.13	-0.19
Variance (%)	14.6	11.1	10.5	6.0	6.0	6.8	6.7	5.9	5.9	5.4
Cumulative Variance (%)	14.6	25.7	36.2	42.2	48.2	55	61.7	67.6	73.5	78.9

Table 4: Table 2. SA Youth Data: Loadings of the First Ten Modified PCs by REM approach. Empty cells have zero loadings.

Discussion

In this work, I have used two approaches to sparse PCA to evaluate its effectiveness in dimensionality reduction and variable selection. The results indicate that the reconstruction error minimization (REM) approach, proposed by Zou, Hastie, Tibshirani (2006), to sparse PCA performs sufficiently better at balancing for sparseness and variance explained by each PC. Its strength is shown by shrinking to zero 40 coefficient loadings, for an affordable loss in explained variance of about 2.9% from the classic PCA. The REM produces the exact results as the PCA when the sparsity penalty term is removed. The REM derives a benefit of identifying important variables, making it a useful benchmark for any potential better method. However, if the goal is concentrated on sparseness more than high variance explained, VM approach is the route to take. The VM approach showed that it can reduce the dimensionality and exclude more irrelevant predictors to a relatively higher degree, but at a cost of explained variance. Moreover, the results show that if a coefficient loading in the classic PCA is close to zero, it will be most likely shrunk to zero by the sparse PCA.

With sparse PCA, I can able to generate PCs which are not contaminated with noise in a such way that the resulting PCs may not deviate from the original data; which is exactly what the the classic PCA suffers greatly from. As a pre-step before applying machine learning techniques, sparse PCA is able to select only the effective variables to construct PCs that can be used to predict the target variable, without the need for us to incorporate variable selection penalty in our regression models. Thus, statistical modelling techniques that do not consider variable selection can leverage on this sparse PCA's strength to improve their robustness.

The future addition of this project will be to use the sparse principal components to build a model that can predict which youth individuals will find work. It is quite useful again to derive the kernel PCA and analyze whether the youth survey datasets are better explained in nonlinear manifold dimensions.

Appendix

Variables	Description	Data Type
Person ID	A unique identifier of all the individuals who participated in the survey	character
Survey Date	Exact date at which the survey was done for each individual	yyyy-mm-dd
Round	The four rounds of 6-month interval survey	numeric
Status	The labour market status of the individual	character
Tenure	(not defined in the data manual)	numeric
Geography	Rural, suburb, or urban ?	character
Province	Provincial residence of the youth individual	character
Matric	Whether or not the youth individual has a matric certificate.	Binary numeric
Degree	Whether or not the youth individual has a degree	Binary numeric
Diploma	Whether or not the youth individual has a diploma	Binary numeric
School Quintile	South Africa's education system is divided into five quintiles based on the socio-economic status of the area surrounding the school	categorical numeric

Table 5: SA Youth: A list of all the original variables before pre-processing of the data.

Variables	Description	Data Type
Math	Mathematics result in matric level	numeric
Mathlit	Mathematical literacy result in matric level	numeric
Additionallang	Additional language result in matric level	numeric
Homelang	Home language result in matric level	numeric
Science	Physical Science or Life Sciences result in matric level	numeric
Female	Gender identity of the youth individual	binary numeric
Sacitizen	Is the youth individual a South African citizen ?	binary numeric
Birthyear	What year was the youth individual born ?	numeric
Birthmonth	What month was the youth individual born ?	numeric
Target	Is the youth individual employed or unemployed after a year of the baseline survey	binary numeric

Table 6: SA Youth: A list of all the original variables before pre-processing of the data.

References

- d'Aspremont, A., Bach, F. and El Ghaoui, L., 2008. Optimal Solutions for Sparse Principal Component Analysis. Journal of Machine Learning Research, 9(7).
- Croux, C., Filzmoser, P. and Fritz, H., 2014. Robust sparse principal component analysis. Quality control and applied statistics, 59(3), pp.231-234.
- Hsu, Y.L., Huang, P.Y. and Chen, D.T., 2014. Sparse principal component analysis in cancer research. Translational cancer research, 3(3), p.182.
- Stone, D., 2020. An Exploration of Alternative Features in Micro-Finance Loan Default Prediction Models. Department of Statistical Sciences, University of Cape Town (UCT).
- Zou, H., Hastie, T. and Tibshirani, R., 2006. Sparse principal component analysis. Journal of computational and graphical statistics, 15(2), pp.265-286.