Dimension Reduction for Prostate Cancer Prediction

Jared Lakhani

20 March 2024

University of Cape Town

Introduction

- According to WHO - cancer is one of the leading causes of death. Nearly ten million deaths in 2020 (or one in every six deaths) [1], with prostate cancer being the most common cancer in South African men [2].
- Predicting cancer in its early stages significantly improves the chances of successful treatment and recovery.
- Pressing need for reliable cancer prediction tools capable of accurately classifying tumors as either malignant (harmful) or benign (non-harmful).
- Use of all features in data set for model building = curse of dimensionality. Need for dimension reduction arises.

Aim

- Whether an ideal dimension reduction technique exists for tumor prognosis model building.
- Best possible parameters for each technique.
- Assessment on classification algorithms used.
- Hypothesise: model building on a reduced feature set results in adequate predictive performance.

Prostate Tumor Dataset

- From digitized images of a cell nuclei present in a fine needle aspiration of a prostate tumor mass.
- 100 Instances with 8 features: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Symmetry, Fractal Dimension
- Response variable: Diagnosis Result (0 for benign and 1 for malignant)

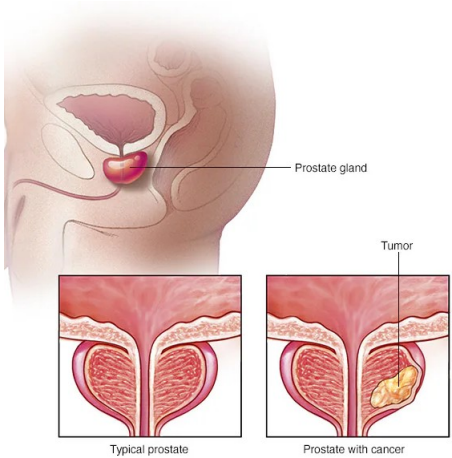


Figure: Prostate Tumor [3]

Prognosis Model

- Data split into training/test set.
- Dimension reduction: Linear, kernel, sparse, robust PCA. Principal curves and cluster analysis.
- Reduced number of features used to train model: SVM-RBF Kernel, Logistic Regression, Naive Bayes.
- Model used on test set to classify tumor: malignant/benign.
- Compared with response variable in test set: Accuracy, Sensitivity, Specificity.

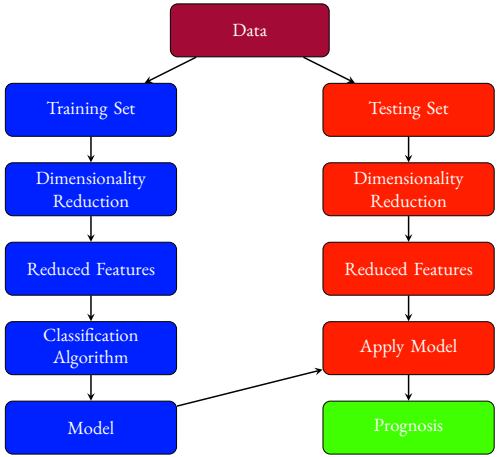
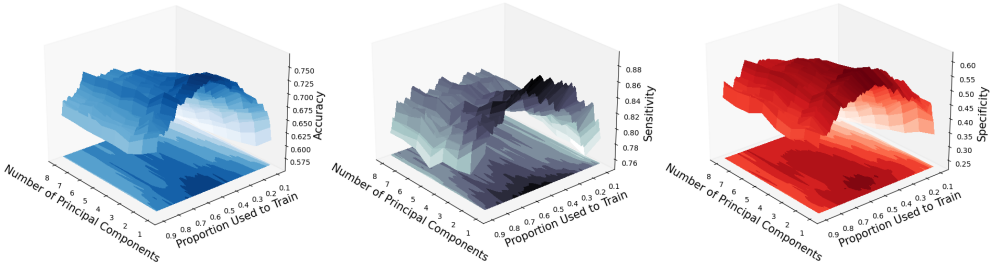


Figure: Flow Diagram of Prognosis Model

Principal Component Analysis

- Bedrock for dimensionality reduction.
- Clear pattern: worsened prediction performance as number of PCs used are increased.
- Greatest performance measures: when the train-test split $\approx 55/45$.



(a) Accuracy vs Number of PCs and Train-Test Split

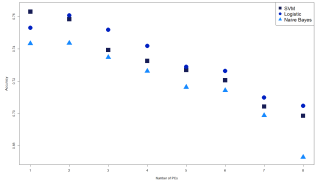
(b) Sensitivity vs Number of PCs and Train-Test Split

(c) Specificity vs Number of PCs and Train-Test Split

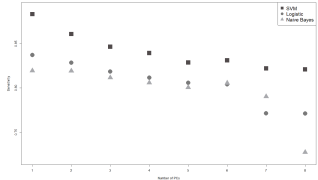
Figure: Performance Measures (Mean) vs Number of PCs and Train-Test Split using SVM (100 Trials)

Principal Component Analysis

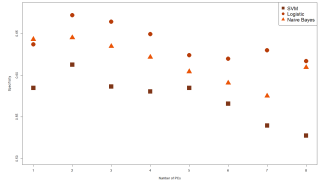
- No single best classification method.
- First PC captures $\approx 37.5\%$ of total data variation.
- More principal components used = greater the amount of total variation of the data explained = a greater quality model?
- Forego possibility that more principal components may increase amount of redundant/irrelevant information in model. Overfitting: model learned to fit noise in data and not underlying patterns.



(a) Accuracy vs Number of PCs



(b) Sensitivity vs Number of PCs

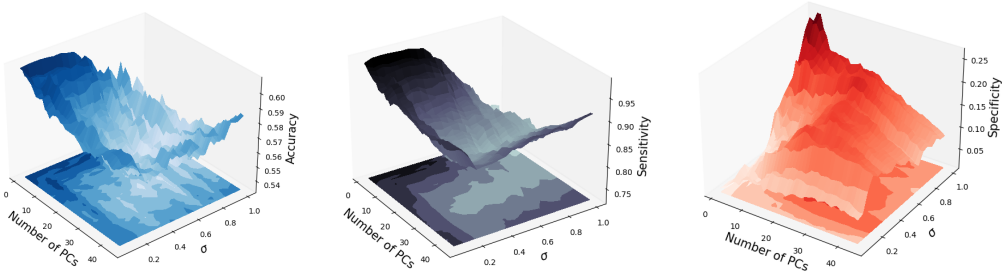


(c) Specificity vs Number of PCs

Figure: Performance Measures (Mean) vs Number of PCs for 55/45 Train-Test Split (1000 Trials)

Kernel Principal Component Analysis - RBF Kernel

- Underlying structure of data may be nonlinear. KPCA can capture complex nonlinear relationships amongst variables.
- No overarching common trend: no ideal σ nor number of PCs to use.

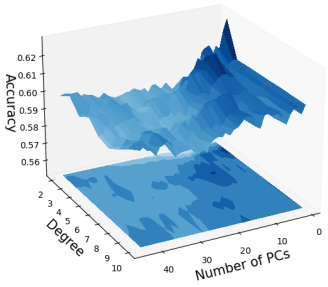


(a) Accuracy vs σ and Number of PCs (b) Sensitivity vs σ and Number of PCs (c) Specificity vs σ and Number of PCs

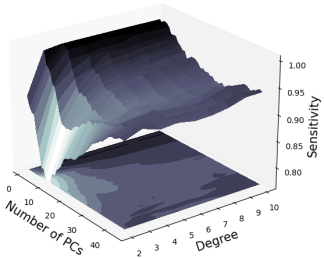
Figure: Performance Measures (Mean) vs σ and Number of PCs using SVM and 55/45 Train-Test Split (100 Trials)

Kernel Principal Component Analysis - Polynomial Kernel

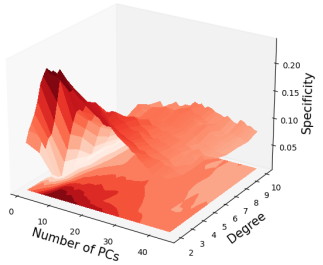
- No clear pattern emerges: no ideal polynomial degree nor number of PCs to use.



(a) Accuracy vs Degrees and Number of PCs



(b) Sensitivity vs Degree and Number of PCs

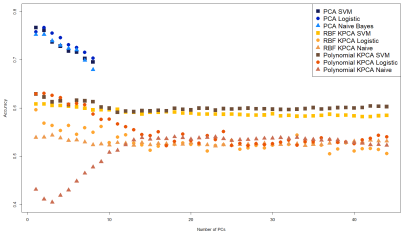


(c) Specificity vs Degree and Number of PCs

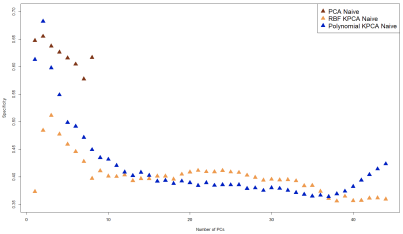
Figure: Performance Measures (Mean) vs Degree and Number of PCs using SVM and 55/45 Train-Test Split (100 Trials)

Kernel Principal Component Analysis

- Accuracy: Linear PCA always out-performed KPCA.
- Specificity: Linear PCA out-performed KPCA (except when quadratic polynomial with 2 PCs used with Naive Bayes)



(a) Accuracy vs Number of PCs for Linear PCA, RBF KPCA ($\sigma = 0.05$) and Polynomial KPCA (degree = 3) for SVM, Logistic Regression and Naive Bayes (100 Trials)



(b) Specificity vs Number of PCs for Linear PCA, RBF KPCA ($\sigma = 0.3$) and Polynomial KPCA (degree = 2) for Naive Bayes (100 Trials)

Kernel Principal Component Analysis

- Can obtain greater sensitivities using Kernel PCA, mostly through use of SVM as the classification method.
- Must assess model's predictive quality by taking all performance measures into account.
- High sensitivity (with low accuracy and specificity) suggests model predicted most tumors malignant - which although would eliminate large number of FNs, is not a prudent choice for a model.

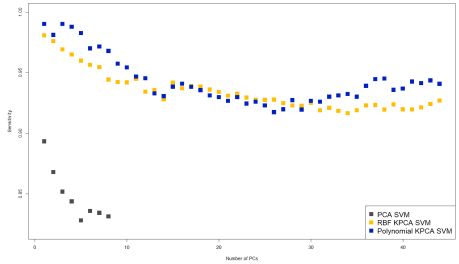


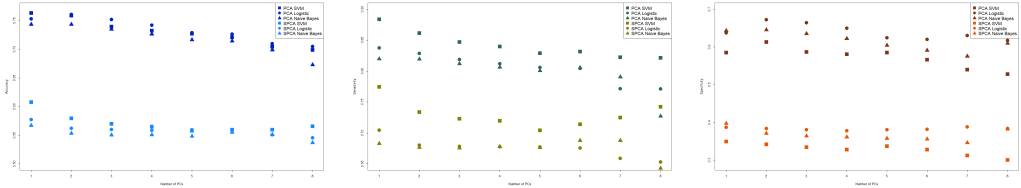
Figure: Sensitivity vs Number of PCs for Linear PCA, RBF KPCA ($\sigma = 0.05$) and Polynomial KPCA (degree = 7) for SVM (100 Trials)

Kernel Principal Component Analysis - Conclusion

- Conclude Kernel PCA overall inferior to linear PCA (only had greater sensitivities).
- KPCA can be prone to overfitting: since it uses a high-dimensional feature space, may capture spurious patterns in the data.
- Designed to capture nonlinear relationships in data, if data is inherently linear or contains only weak nonlinearities, may perform worse.

Sparse Principal Component Analysis

- Encourages sparsity in the loadings of the principal components: can filter out noisy or irrelevant features.
- Same nature to PCA: decreased performance with increasing number of PCs used.



(a) Accuracy vs Number of PCs (b) Sensitivity vs Number of PCs (c) Specificity vs Number of PCs

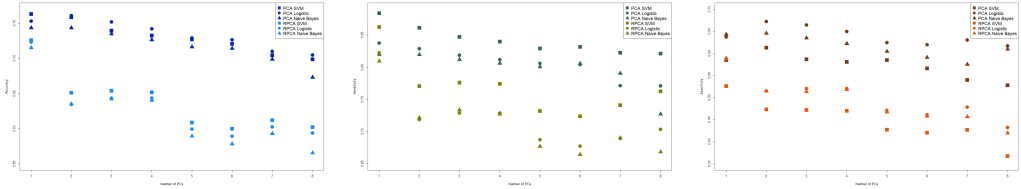
Figure: Performance Measures (Mean) vs Number of PCs for PCA and SPCA for all Classification Methods for 55/45 Train-Test Split (1000 Trials)

Sparse Principal Component Analysis - Conclusion

- Inferior performance relative to PCA with respect to all performance measures (and classification methods)
- Sparse PCA: only a subset of original features contribute significantly to each principal component. This sparsity may lead to a loss of information.
- PCA: considers all features equally in each principal component and may provide a more balanced representation.

Robust Principal Component Analysis

- Combats against outliers (potential "corrupted observations") having a significant effect.
- Performs worse than PCA using default robustness $\alpha = 0.75$.
- Possibly due to loss of information - robust estimators may downweight or exclude observations that are considered outliers. PCs less representative than PCA.



(a) Accuracy vs Number of PCs (b) Sensitivity vs Number of PCs (c) Specificity vs Number of PCs

Figure: Performance Measures (Mean) vs Number of PCs for PCA and RPCA for all Classification Methods for Default Robustness $\alpha = 0.75$ and 55/45 Train-Test Split (100 Trials)

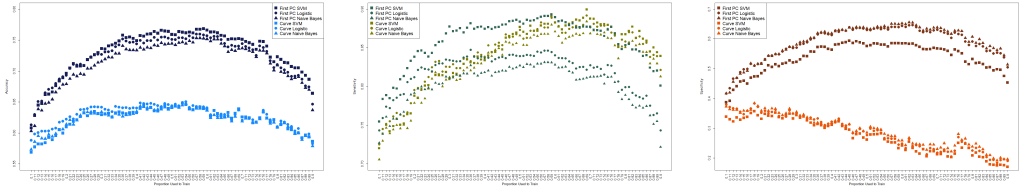
- With maximum robustness $\alpha = 1$ and using the first PC: similar performance measures to PCA - performance measures fall off greatly after this however.

		Accuracy			Sensitivity			Specificity		
PCs		PCA	RPCA$^{(\alpha_{0.75})}$	RPCA$^{(\alpha_1)}$	PCA	RPCA$^{(\alpha_{0.75})}$	RPCA$^{(\alpha_1)}$	PCA	RPCA$^{(\alpha_{0.75})}$	RPCA$^{(\alpha_1)}$
1	SVM	0.767	0.720	**0.771**	0.893	0.868	**0.895**	0.588	0.511	**0.601**
	Logistic	0.758	0.719	**0.767**	0.852	0.825	**0.855**	0.635	0.580	**0.656**
	Naive	0.752	0.706	**0.756**	0.836	0.808	**0.838**	0.647	0.575	**0.659**

Table: Performance Measures (Mean) with Varying Number of PCs and Classification Methods for PCA and RPCA for Default Robustness Parameter $\alpha_{0.75} = 0.75$ and $\alpha_1 = 1$ and 55/45 Train-Test Split (100 Trials)

Principal Curves

- Concluded that solely using the first PC (from PCA) would suffice in training prognosis model.
- Analogous to PCA: reduced feature is the 'projection' of the observation onto the principal curve. Worthwhile comparing first PC (from PCA) to principal curve.
- Using first PC overall superior to using a principal curve.

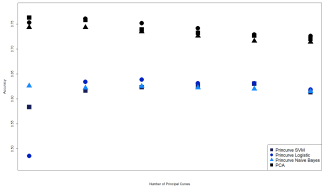


(a) Accuracy vs Train-Test Split (b) Sensitivity vs Train-Test Split (c) Specificity vs Train-Test Split

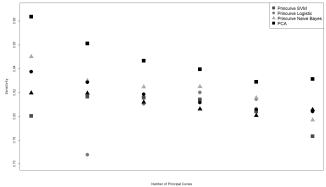
Figure: Performance Measures vs Train-Test Split for Principal Curve and First PC for all Classification Methods (250 Trials)

K-Means + Principal Curves

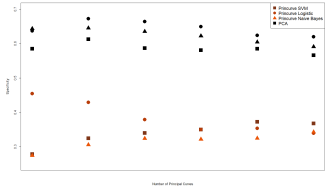
- Hybrid approach: Initially partition observations into clusters. Fit principal curve to each set of observations belonging to a cluster.
- Only improves accuracy.
- Still inferior to PCA.



(a) Accuracy vs Number of Principal Curves



(b) Sensitivity vs Number of Principal Curves



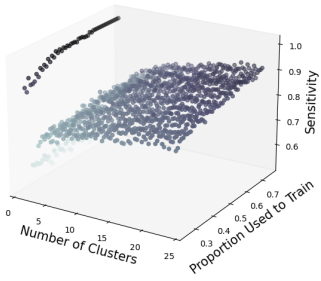
(c) Specificity vs Number of Principal Curves

Figure: Performance Measures vs Number of Principal Curves all Classification Methods (1000 Trials)

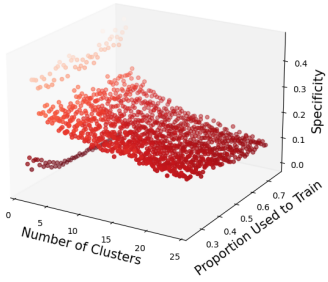
Cluster Analysis

- New feature is the cluster to which the specific observation belongs.
- Using a single cluster: model classifies all tumors as malignant - sensitivity $\approx$ 1 (Only FPs and TPs) and specificity $\approx$ 0 (No TNs).

(a) Accuracy vs Number of Clusters and Train-Test Split



(b) Sensitivity vs Number of Clusters and Train-Test Split

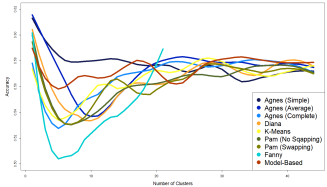


(c) Specificity vs Number of Clusters and Train-Test Split

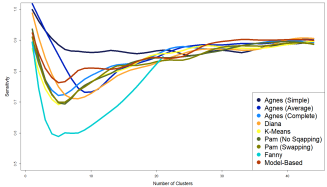
Figure: Performance Measures vs Number of Clusters and Train-Test Split using Naive Bayes and K-Means Clustering (200 Trials)

Cluster Analysis

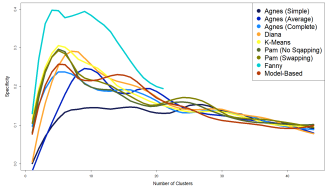
- Similar nature: accuracy and sensitivity decrease to a minimum and increase thereafter (opposite for specificity).
- This min/max value differs for each clustering technique.
- No single best clustering method (nor classification method).
- Regardless, inferior to PCA.



(a) Accuracy vs Number of Clusters



(b) Sensitivity vs Number of Clusters



(c) Specificity vs Number of Clusters

Figure: Performance Measures vs Number of Clusters for Agnes (Single, Average and Complete Linkage), Diana, K-Means, Pam (No Swapping and Swapping), Fanny and Model-Based Clustering Techniques for SVM (100 Trials) with LOESS Smoothing

Conclusion

- Principle Component Analysis: superior technique for building prostate tumor prognosis model.
- Just use first principle component: overall greatest accuracy, sensitivity, specificity.
- Performs well against training model on full feature set.

	Accuracy	Sensitivity	Specificity
All 8 features	81%	89%	70%
First PC	**76%**	**88%**	**59%**

Table: Performance Measures (Mean) of Prognosis Model Trained from Full Feature Set and First PC using SVM-RBF (1000 Trials)

- No single best classification method (Try: tuning cost and γ of RBF Kernel for SVM or other algorithms)

References I

1. World Health Organization. *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer.
2. Foundation, M. *Prostate Cancer*. Accessed: March 19, 2024. https://mensfoundation.co.za/mens-health/prostate-cancer/.
3. Mayo Clinic. *Prostate cancer*. Mayo Clinic. Retrieved March 17, 2024.