

## PCA - Principal Component Analysis

Hernane Braga Pereira - 2014112627

### 1. Introdução

Este relatório tem como objetivo demonstrar o uso da técnica PCA, Principal Component Analysis, aplicada a redução de dimensionalidade de problemas de classificação não linearmente separáveis.

### 2. Base *breastcancer*

Para este exercício foi pedido que classifica-se o problema *breastcancer*, do pacote *mlbench* do R, que é uma base real de detecção de câncer de mama. O dataset possui 9 variáveis de entrada e as classes de saída são se o diagnóstico detectou células benignas ou malignas, representadas pelos valores 1 e -1, respectivamente.

Para escolher quais características da base de dados seriam utilizadas, foi aplicado o método PCA e os valores dos autovalores e sua variância foram analisados e podem ser vistos nas figura 1 e 2.

#### Autovalores das Características

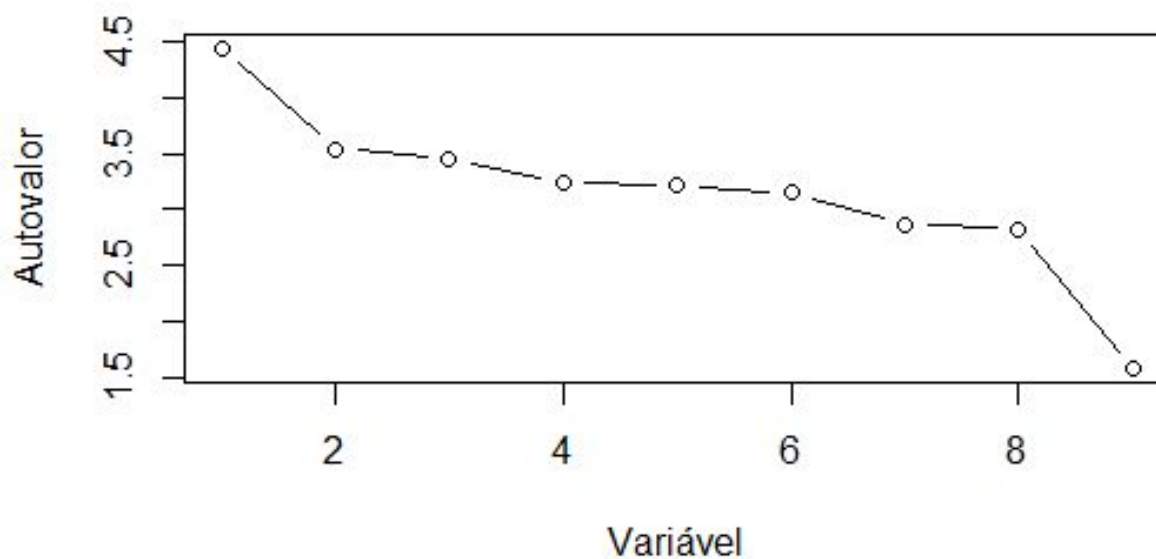


Figura 1. Autovalores das características

#### Variância das Características

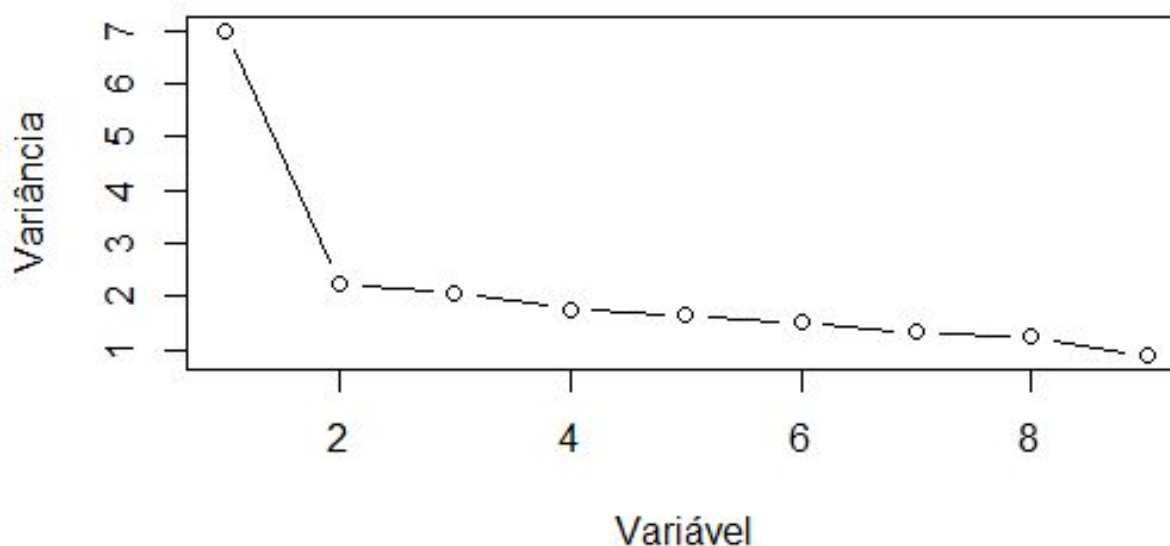


Figura 2. Variância das características

Após esta análise escolheu-se as primeiras 3 características para serem as entradas do problema de classificação, que foi resolvido utilizando o método SVM, usando a função *ksvm* da biblioteca *kernelab* do R. No método, a função gaussiana foi usada como kernel, e os parâmetros de entrada foram  $\sigma = 0.5$  e parâmetro  $C = 5$ . Os dados foram separados em 10 folds para validação cruzada e 90% dos dados foram usados para treino e 10% para testes. Os resultados de acurácia de cada classe podem ser vistos no quadro 1.

Teste	Benigno	Maligno	Total
1	96	97,7	97
2	96,2	95,2	95,5
3	88	100	95,5
4	96,2	100	98,5
5	95,7	95,6	95,5
6	100	95,7	97,1
7	96,3	100	98,5
8	91,3	97,8	95,6
9	100	100	100
10	92,9	97,6	95,7
Média	95,26	97,96	96,89

Quadro 1. Acurácia obtidas para 3 características

### 3. Referências

[1] PCA, Notas de aula, outubro de 2019.  
[2] SVM, Notas de aula, outubro de 2019.