

Redução de Dimensionalidade usando uma modificação do algoritmo Autofaces Fracionário

Emanuel D. Ferreira dos Santos, Tiago Buarque de Carvalho

6 de novembro de 2018

Resumo

Este trabalho visa expor os resultados da modificação do algoritmo Autofaces Fracionário. O método proposto funciona substituindo o parâmetro escalar r , por um vetor de tamanho variável R . A fim de encontrar os valores do vetor R que maximizassem o valor da acurácia, escolhemos usar um Algoritmo Genético como meta-heurística capaz de nos fornecer esse resultado. Nos experimentos, comparamos o método proposto com o Autofaces Fracionário, aplicando ambos em seis bases de dados reais e calculando métricas de avaliação usando o classificador 1-NN.

palavras-chave: redução de dimensionalidade, PCA, Algoritmo Genético, Autofaces Fracionário.

1 Introdução

No contexto de redução de dimensionalidade, o termo *dimensionalidade* se refere a quantidade de variáveis aleatórias que um determinado conjunto de dados possui. Assim, se uma amostra x_i é definida por

$$x_i = [\alpha_1, \alpha_2, \dots, \alpha_n]$$

dizemos que a amostra x_i possui dimensionalidade n , sendo que $n \in \mathbb{N}^*$.

A redução de dimensionalidade acontece quando conseguimos diminuir o número de atributos do conjunto original de modo que ainda tenhamos uma versão que seja representativa dos dados originais.

Para o Aprendizado de Máquina, a importância da redução de dimensionalidade consiste em poder submeter para os algoritmos apenas os atributos que sejam mais relevantes no que se refere ao manuseio dos dados.

2 Algoritmos Relacionados

Muitos métodos têm sido propostos a fim de melhorar a qualidade dos algoritmos de redução de dimensionalidade. Alguns deles derivam do método algébrico PCA, o qual é conhecido como o mais bem estabelecido na literatura. Nessa seção falaremos sobre o funcionamento do PCA e três métodos derivados dele: FPCA, Autofaces e Autofaces Fracionário.

2.1 PCA

PCA (Principal Component Analysis, em português Análise dos Componentes Principais) é uma técnica de redução de dimensionalidade que ao receber como entrada um dataset D com dimensão m , produz uma projeção de D no espaço de dimensionalidade m' , sendo $m' < m$.

Seja D uma matriz que contém os dados a serem reduzidos. Temos que a matriz D , contendo n amostras é definida como

$$D = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

Cada amostra x_i contém m variáveis, as quais chamamos de atributos. Desta forma temos

$$x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$$

O vetor médio \bar{x} é do conjunto de dados é obtido através da fórmula:

$$\bar{x} = \frac{1}{n} \sum x_i$$

Subtraindo o vetor médio de cada amostra de D , teremos uma versão dos dados centralizada na origem:

$$D' = \begin{bmatrix} (x_1 - \bar{x}) \\ (x_2 - \bar{x}) \\ (x_3 - \bar{x}) \\ \vdots \\ (x_n - \bar{x}) \end{bmatrix}$$

A partir de D' é possível obter a matriz de covariância dos dados. Essa matriz contém o valor da covariância entre cada par de atributos, bem como a variância de cada atributo na diagonal principal. Ela pode ser obtida através de

$$Cov(D') = \frac{1}{n} D'^T D'$$

A matriz $Cov(D')$ será de ordem $m \times m$.

Cada coluna da matriz $Cov(D')$ é um autovetor. Dessa forma temos que

$$Cov(D') = [\xi_1, \xi_2, \dots, \xi_m]$$

Onde para cada autovetor ξ_i existe um autovalor λ_i associado, o qual representa a variância daquele atributo.

NÃO CONCLUIDO

2.2 PCA Fracionário

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor. Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.

2.3 Autofaces

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor. Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.

2.4 Autofaces Fracionário

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor. Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.

3 Algoritmo Genético

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor.

Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.

4 Método Proposto

O método proposto consiste em modificar o Autoface Fracionário por meio da aplicação de um vetor R no lugar da escalar r . Cada elemento de R seria aplicado sobre uma fração da amostra. Para o caso específico de imagens, um vetor R de comprimento n implicaria que a amostra seria dividida em uma matriz $\sqrt{n}x\sqrt{n}$. Assim teríamos uma variação maior do parâmetro r original do Autofaces Fracionário sobre a amostra de entrada.

Para a obtenção do vetor R , foi usado um Algoritmo Genético com o objetivo de se conseguir bons valores para o vetor e que por consequência melhorasse a qualidade da redução de dimensionalidade e por conseguinte a acurácia do algoritmo de classificação utilizado.

5 Experimentos

5.1 Metodologia Adotada

A metodologia dos experimentos realizados para calcular os valores do vetor R pode ser conferida no fluxograma da figura 1. Após a obtenção do vetor R , um outro algoritmo é executado a fim de comparar os desempenhos do Autofaces Fracionário e do Autofaces Fracionário proposto. Esse algoritmo pode ser observado na figura 2.

Os valores de componentes testados em cada execução do Algoritmo Genético foram os seguintes:

$$c = [1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70]$$

Assim, cada experimento executou um Algoritmo Genético 15 vezes, uma para cada número de componentes. Além disso, cada execução do AG resultava em um vetor R para algum número de componentes considerado no momento.

5.2 Resultados

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui.

6 Conclusões

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo li-

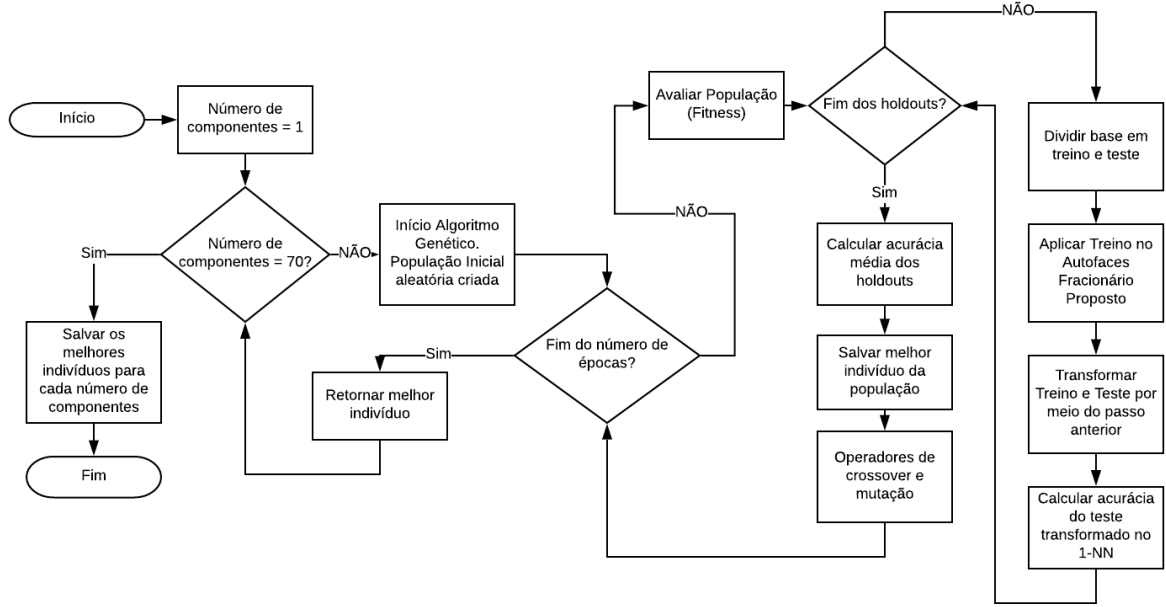


Figura 1: Digrama do algoritmo genético utilizado para obter valores de R .

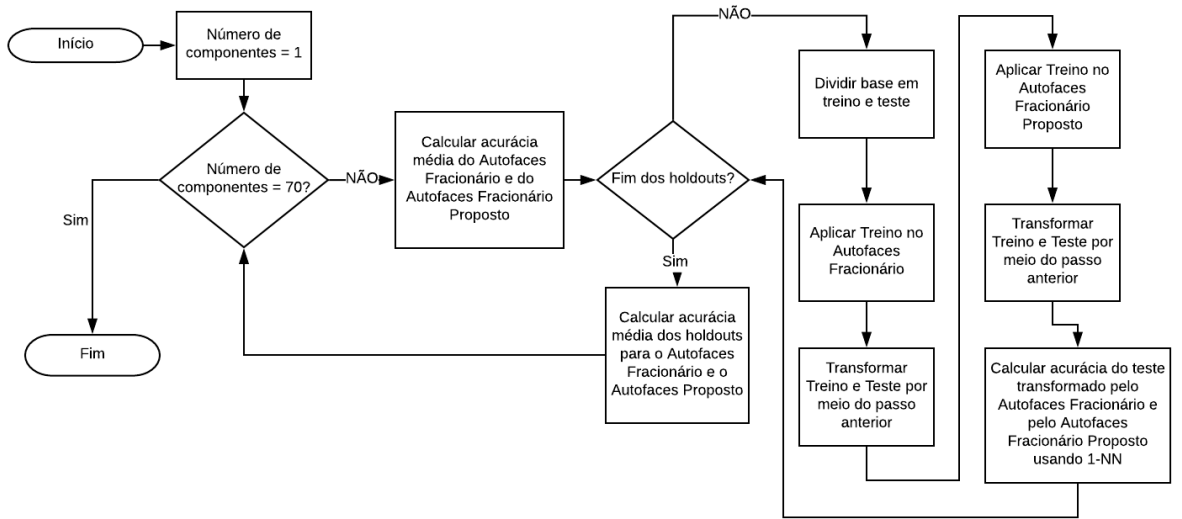


Figura 2: Digrama do algoritmo que compara o desempenho entre o Autofaces Fracionário e o proposto.

bero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor. Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam

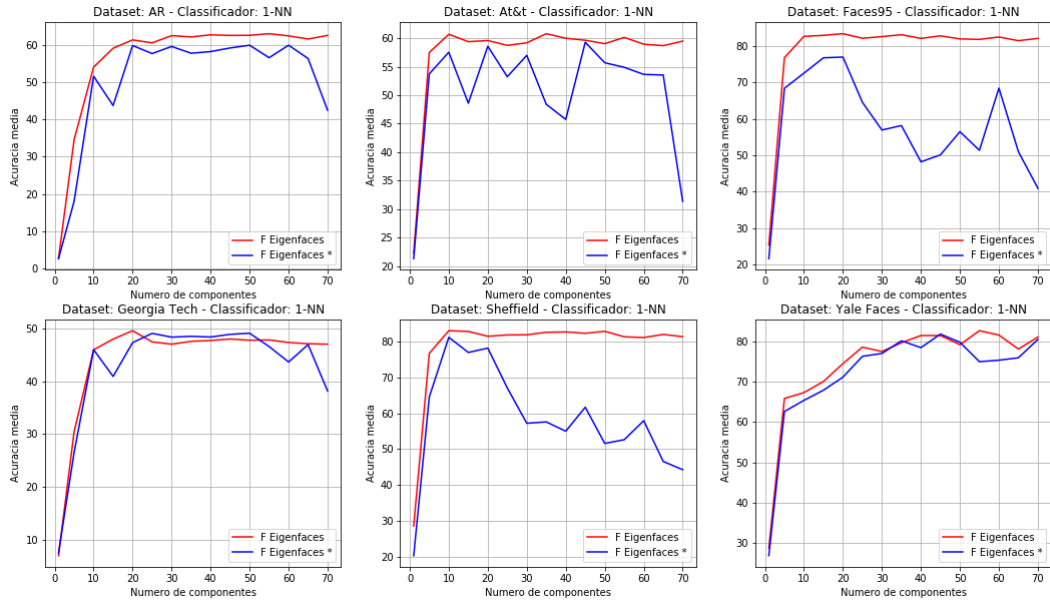


Figura 3: Resultados para os seis datasets

porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.

7 Referências

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent porta, lorem non faucibus condimentum, orci orci condimentum lectus, sit amet sollicitudin justo libero non dui. Sed sollicitudin leo at urna dapibus cursus. Vestibulum nibh dui, tristique vitae mollis tristique, sagittis pretium justo. Nulla faucibus suscipit dolor eu semper. Nam nisl lacus, tempor quis nisi vitae, dictum sollicitudin dolor. Vivamus quis elit consequat, ullamcorper ligula non, ultricies enim. Proin dictum leo eget tempor placerat. Cras dapibus eget lorem sed malesuada. Mauris lobortis iaculis purus, eu varius neque lacinia et. Vivamus gravida nisl in pellentesque molestie. Aliquam sit amet urna ut nulla faucibus eleifend sit amet sed nibh. Nam porta leo sed enim interdum posuere. Duis purus diam, volutpat vel luctus sit amet, venenatis at lacus.