

# Análise de Componentes Principais (PCA)

Técnicas de Redução de Dimensionalidade

Aurélio Figueiredo  
[aurelio.figueiredo@foa.org.br](mailto:aurelio.figueiredo@foa.org.br)

# Objetivos

Entender o que são métodos de redução de dimensionalidade;

Descrever o método de Análise de Componentes Principais;

O que pode ser feito quando se tem um conjunto de exemplos que possua muitas variáveis de entrada, de forma a tentar contornar a chamada Maldição de Dimensionalidade?

# Redução de Dimensionalidade

A redução de dimensionalidade é a transformação de dados de um espaço de alta dimensão em um de baixa dimensão de forma que a nova representação retenha o máximo possível das propriedades significativas dos dados originais, idealmente perto de sua dimensão intrínseca;

Trabalhar em espaços de grande dimensão pode ser indesejável por muitos motivos: os dados brutos costumam ser esparsos como consequência da maldição da dimensionalidade, e a análise dos dados costuma ser computacionalmente intratável;

A redução da dimensionalidade é comum em campos que lidam com um grande número de observações e/ou um grande número de variáveis, como em Aprendizado de Máquina.

# Análise de Componentes Principais (PCA)

A Análise de Componentes Principais ou Principal Component Analysis (PCA) é um procedimento matemático que utiliza uma transformação ortogonal (ortogonalização de vetores) para converter um conjunto de variáveis possivelmente correlacionadas num conjunto de valores de variáveis linearmente não correlacionadas chamadas de componentes principais;

O número de componentes principais é sempre menor ou igual ao número de variáveis originais. O PCA é sensível à escala relativa das variáveis originais. Dependendo da área de aplicação, o PCA é também conhecido como transformada de Karhunen-Loève (KLT) discreta, transformada de Hotelling ou decomposição ortogonal própria (POD).

# Análise de Componentes Principais (PCA)

Portanto, a análise de componentes principais (PCA) é um método que tem por finalidade básica, análise dos dados de entrada usados visando sua redução, eliminação de sobreposições e a escolha das formas mais representativas dos dados a partir de combinações lineares das variáveis originais.

Análise de Componentes Principais (PCA) é um dos métodos estatísticos de múltiplas variáveis mais simples, e mais utilizados que existem. O PCA é considerado a transformação linear ótima, dentre as transformadas de datasets de alta dimensionalidade, sendo muito utilizada pela comunidade de Aprendizado de Máquina.

# Análise de Componentes Principais (PCA)

O PCA é matematicamente definido como uma transformação linear ortogonal que transforma os dados para um novo sistema de coordenadas, de forma que a maior variância por qualquer projeção dos dados fica ao longo da primeira coordenada (o chamado **componente principal**), a segunda maior variância fica ao longo da segunda coordenada, e assim por diante;

Dado um conjunto de pontos no espaço euclidiano, o primeiro componente principal corresponde a uma linha que passa através da média multidimensional e minimiza a soma dos quadrados das distâncias dos pontos à linha;

O segundo componente principal corresponde ao mesmo conceito, depois de subtrair-se toda a correlação com o primeiro componente principal dos pontos, e assim por diante;

# Análise de Componentes Principais (PCA)

O PCA essencialmente rotaciona o conjunto de pontos em torno da média de forma a alinhá-los com os componentes principais. Isto move o máximo possível de variância (usando uma transformação ortogonal) a algumas das primeiras dimensões;

Os valores nas dimensões restantes, portanto, tendem a ser pequenos e podem ser descartados com o mínimo de perda de informação. O PCA é comumente utilizado dessa maneira para redução de dimensionalidade;

O PCA tem a distinção de ser a melhor transformação ortogonal para manter o subespaço que tem a maior "variância".

# Bibliografia Básica

NORVIG, Peter; RUSSELL, Stuart Jonathan. Inteligência Artificial. Rio de Janeiro: Elsevier, 2004.

COPPIN, Ben. Inteligência Artificial. Rio de Janeiro: LTC, 2012.

COSTA, Ernesto. Inteligência Artificial: Fundamentos e Aplicações. 3. ed. Lisboa: FCA, 2008.

ROSA, João Luis Garcia. Fundamentos da Inteligência Artificial. Rio de Janeiro: LTC, 2011.



# Bibliografia Complementar

FERNANDES, Anita Maria da Rocha. Inteligência Artificial: Noções Gerais. Florianópolis: Visual Books, 2008.

FINGER, Marcelo; MELO, Ana Cristina Vieira de; SILVA, Flávio Soares Corrêa da. Lógica para Computação. São Paulo: Thomson, 2006.

LÉVY, Pierre. As Tecnologias da Inteligência: O Futuro do Pensamento na Era da Informática. São Paulo: Editora 34, 2006.