

Métodos para Redução de Dimensionalidade

Cesar de Oliveira F. Silva

Setembro / 2026

Roteiro

- 1 Motivação e Intuição
- 2 Seleção vs. Extração de Atributos
- 3 PCA – Análise de Componentes Principais
- 4 Resumo

Por que reduzir dimensionalidade?

- Muitas aplicações atuais geram **dados com centenas ou milhares de atributos**: dados genômicos, imagens de satélite, sensores ambientais, etc.
- Nem todos esses atributos são úteis:
 - muitos podem ser **irrelevantes** ou redundantes;
 - a **dimensão intrínseca** do problema costuma ser bem menor.
- Objetivos principais:
 - melhorar a performance de algoritmos de aprendizado de máquina;
 - simplificar modelos e reduzir custo computacional;
 - facilitar visualização e interpretação dos resultados.

A maldição da dimensionalidade

- À medida que o número de dimensões cresce, os dados ficam **progressivamente esparsos** no espaço.
- Distâncias entre pontos tendem a ficar todas parecidas:
 - a diferença entre a distância mínima e máxima entre pares de pontos diminui;
 - fica difícil dizer quem é “próximo” de quem.
- Analogia:

Procurar uma agulha...

Em 2D é como procurar uma agulha em um campo de futebol. Em 100D é como o mesmo campo, mas com volume de uma galáxia: os dados “somem” no vazio.

Motivações práticas

- **Visualização:** projetar dados de alta dimensão em 2D/3D para entender padrões, clusters e outliers.
- **Compressão de dados:** guardar projeções em poucas dimensões em vez de todas as variáveis originais.
- **Remoção de ruído:** dimensões com baixa variância ou muito ruído podem ser descartadas, melhorando a acurácia de modelos.
- **Entendimento do problema:** estudar relações entre atributos antes de treinar modelos complexos.

Dois caminhos: Seleção e Extração

Seleção de atributos

Escolher um **subconjunto ótimo** de atributos originais de acordo com uma função objetivo (ex.: acurácia, custo, interpretabilidade).

Extração de atributos

Construir **novas dimensões** como combinações dos atributos originais, preservando o máximo de informação possível.

- Em geral, a seleção é mais fácil de interpretar (ainda estamos falando dos atributos originais).
- Extração tende a ser mais poderosa para compressão, mas os novos eixos podem ser difíceis de “explicar”.

Seleção de atributos

- Ideia geral:
 - encontrar o **menor conjunto de atributos** que ainda mantém bom desempenho preditivo;
 - remover atributos redundantes ou irrelevantes.
- Objetivos:
 - reduzir dimensionalidade e ruído;
 - aumentar velocidade de treinamento e inferência;
 - melhorar interpretabilidade dos modelos.
- Exemplos:
 - filtros estatísticos (correlação, χ^2 , informação mútua);
 - métodos *wrapper* (busca guiada por desempenho);
 - métodos *embedded* (Lasso, árvores, etc.).

Extração de atributos

- Em vez de escolher alguns atributos, criamos **novas dimensões** como combinação linear dos atributos originais.
- Em geral, usamos apenas os atributos numéricos (vetores em $\mathbb{R}^d$), muitas vezes sem utilizar a variável classe.
- Objetivo:
 - projetar os dados em um espaço de menor dimensão **minimizando o erro de reconstrução** ou **preservando distâncias**.
- Exemplos:
 - PCA (Principal Component Analysis);
 - Projeção Aleatória (Random Projection);
 - MDS (Multidimensional Scaling);
 - SVD / LSI (Latent Semantic Indexing).

Seleção vs Extração: comparação

- **Seleção**

- mantém apenas alguns atributos originais;
- resultado mais fácil de explicar;
- pode perder padrões que só aparecem em combinações de atributos.

- **Extração**

- usa *todos* os atributos para gerar novos eixos;
- novos atributos geralmente são combinações lineares;
- maior poder de compressão, mas menos interpretável.

- Na prática:

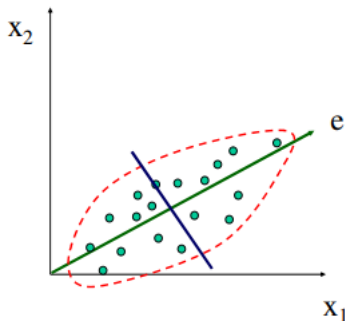
Combinar é comum

Muitas vezes aplicamos **seleção** primeiro (remover lixo grosseiro) e depois **extração** (ex.: PCA) para compactar a informação.

PCA: ideia geral

- PCA transforma variáveis **correlacionadas** em um novo conjunto de variáveis **não correlacionadas** chamadas de **componentes principais**.
- Cada componente é um eixo no espaço dos dados que:
 - captura o máximo possível de **variância**;
 - é ortogonal aos demais componentes.
- Ao usar apenas as primeiras componentes, encontramos uma **aproximação de baixa dimensão** que preserva a maior parte da informação dos dados originais.

PCA: intuição geométrica



- Os pontos verdes são dados 2D (x_1, x_2).
- A linha verde e é a **primeira componente principal**: direção que melhor “alinha” com a nuvem de pontos.
- A linha azul, ortogonal à primeira, captura a variação restante.

Metáfora

PCA: definição matemática (resumida)

- Considere dados centrados em média: matriz $X \in \mathbb{R}^{n \times d}$.
- Matriz de covariância:

$$S = \frac{1}{n-1} X^T X.$$

- PCA busca autovetores $\mathbf{u}_1, \dots, \mathbf{u}_d$ de S :

$$S\mathbf{u}_k = \lambda_k \mathbf{u}_k,$$

onde λ_k são os autovalores (variância explicada).

- As primeiras componentes principais são os autovetores associados aos maiores autovalores.

PCA: erro de reconstrução e ortogonalidade

- As componentes principais são vetores **ortogonais** que formam uma nova base para o espaço dos dados.
- Se usamos apenas as primeiras k componentes ($k < d$):
 - projetamos cada ponto x em um subespaço de dimensão k ;
 - reconstruímos uma aproximação $\tilde{x}$ dos dados originais.
- PCA escolhe os vetores que **minimizam o erro quadrático médio (RMS)**

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|x_i - \tilde{x}_i\|^2}.$$

- Ou seja, entre todas as projeções de dimensão k , PCA é a que “perde menos informação” em média.

PCA: passos práticos

- 1 **Normalizar** os dados:
 - centralizar em média (subtrair a média de cada atributo);
 - opcionalmente escalar pela variância (z-score).
- 2 Calcular a **matriz de covariância** S .
- 3 Obter autovalores λ_k e autovetores $\mathbf{u}_k$ de S .
- 4 Ordenar componentes em ordem decrescente de λ_k .
- 5 Escolher um número k de componentes e projetar os dados:

$$Z = XU_k,$$

onde U_k contém as k primeiras colunas de U .

Quantas componentes escolher?

- Cada autovalor λ_k indica quanta **variância** é explicada pela componente k .
- A proporção de variância explicada pelas primeiras k componentes é:

$$\text{PVE}(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^d \lambda_j}.$$

- Regra prática:
 - escolher o menor k tal que $\text{PVE}(k)$ esteja em torno de **80–90%**;
 - observar o **scree plot** (gráfico dos autovalores) e procurar o “cotovelo”.

PCA e descarte de atributos

- Dados: p variáveis originais.
- Executar PCA sobre a **matriz de correlação** (ou covariância).
- Obter p componentes principais, com seus autovalores.
- Ideia:
 - componentes com **autovalores baixos** explicam pouca variância;
 - a variável que mais “domina” uma componente pouco importante tende a ser **redundante ou pouco informativa**.

Algoritmo para descarte de variáveis via PCA

- 1 Executar PCA e ordenar os autovalores em ordem crescente.
- 2 Para cada componente de menor autovalor:
 - identificar o **coeficiente de maior módulo** (em valor absoluto) no autovetor correspondente;
 - descartar a variável associada a esse coeficiente.
- 3 Repetir o processo até descartar $(p - k)$ variáveis, preservando apenas k variáveis “mais importantes”.

Princípio

Componentes com baixo autovalor são menos relevantes; se uma variável domina uma componente pouco relevante, é candidata natural ao descarte.

Outros métodos de redução de dimensionalidade

- **Random Projection**

- projeta dados em um subespaço gerado por vetores aleatórios;
- rápido, simples e preserva aproximadamente distâncias.

- **Multidimensional Scaling (MDS)**

- recebe como entrada as distâncias entre pontos;
- encontra uma configuração em baixa dimensão que preserva tais distâncias.

- **SVD / LSI**

- decomposição em valores singulares (muito usada em textos, recomendação, etc.);
- semelhante a PCA aplicada a matrizes esparsas termo–documento.

Resumo e mensagens finais

- Alta dimensionalidade complica: **distâncias perdem significado**, modelos ficam lentos e difíceis de interpretar.
- Redução de dimensionalidade ajuda a:
 - melhorar performance e generalização;
 - simplificar modelos;
 - facilitar visualização e entendimento.
- Dois caminhos complementares:
 - **Seleção de atributos**;
 - **Extração de atributos** (PCA e outros).
- PCA é o método clássico: projeta dados em poucas dimensões preservando a maior parte da variância.