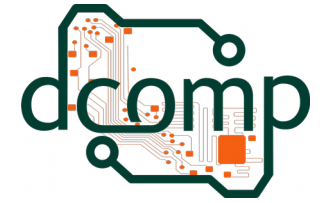




Universidade Federal do Espírito Santo  
Centro de Ciências Agrárias – CCENS UFES  
Departamento de Computação



# Vizinhos mais próximos

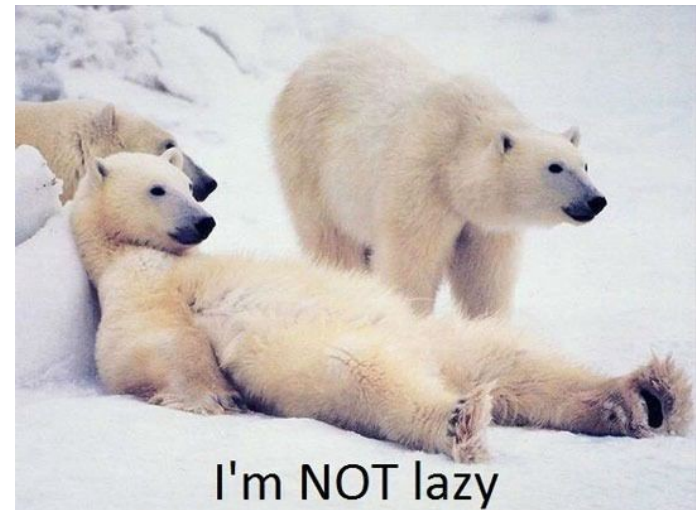
## **Inteligência Artificial**

Site: <http://jeiks.net>

E-mail: [jacsonrcsilva@gmail.com](mailto:jacsonrcsilva@gmail.com)

# Modelos não paramétricos

- Modelos paramétricos:
  - Utilizam dados de treinamento para estimar um conjunto fixo de parâmetros  $w$ .
  - Ao definir a hipótese  $h_w(x)$ , pode-se descartar os dados de treinamentos, pois eles estão resumidos por  $w$ .
  - A quantidade de parâmetros necessários são sempre iguais, mesmo com mais ou menos dados de treinamento.
- Modelo não paramétrico:
  - Não podem ser caracterizados por um conjunto limitado de parâmetros.
  - Cada hipótese que geramos simplesmente mantém dentro de si todos os exemplos de treinamento e os usa para prever o próximo exemplo.
  - Tal grupo de hipótese será não paramétrico:
    - O número efetivo de parâmetros é ilimitado,
    - Cresce com o número de exemplos.
  - Essa abordagem é chamada de:
    - Aprendizagem baseada em exemplos; ou
    - Aprendizagem baseada em memória; ou
    - Aprendizado *lazy* (preguiçoso).



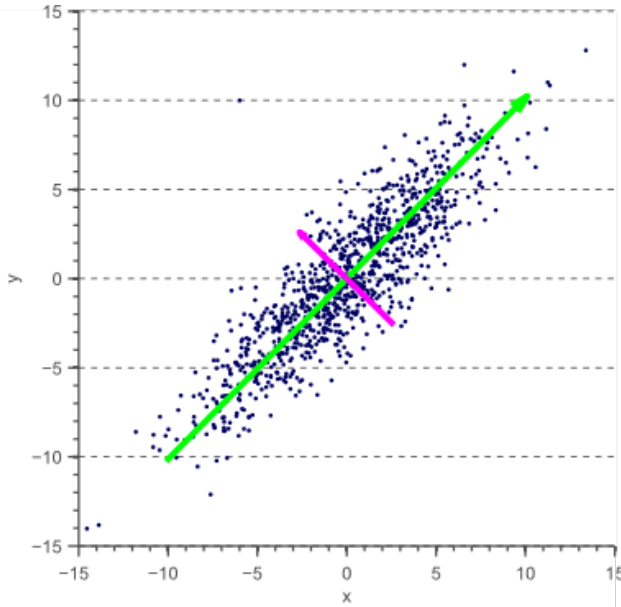
I'm NOT lazy

I'm HIGHLY motivated to do NOTHING

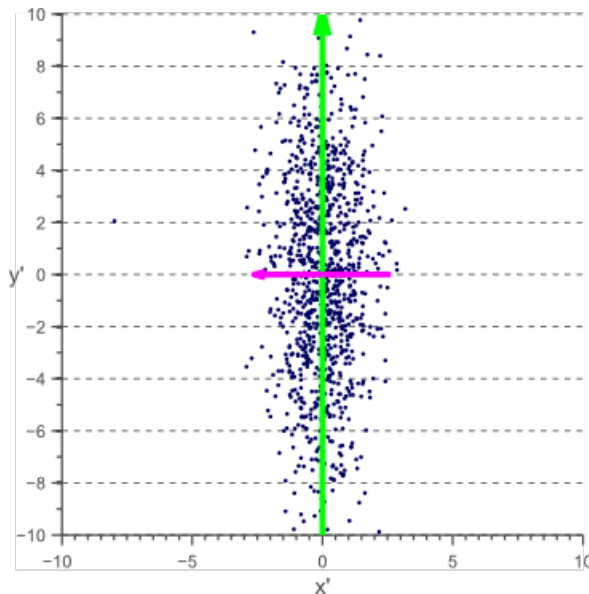
# Dimensionalidade de problemas

- Espaços de dimensão baixa e com abundância de dados:
  - Os vizinhos mais próximos funcionam muito bem.
  - Pode-se ter dados suficientes nas proximidades para obter uma boa resposta.
- Número de dimensões alto:
  - os vizinhos mais próximos geralmente não estão muito próximos.
  - Maldição da dimensionalidade.
- Uma alternativa para trabalhar com os atributos:
  - PCA: Análise de Componentes Principais.

Encontram-se seus autovetores e autovalores:

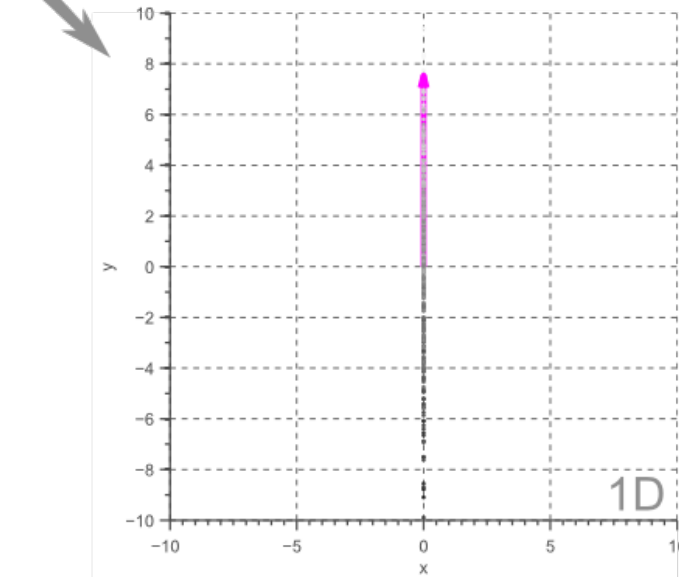
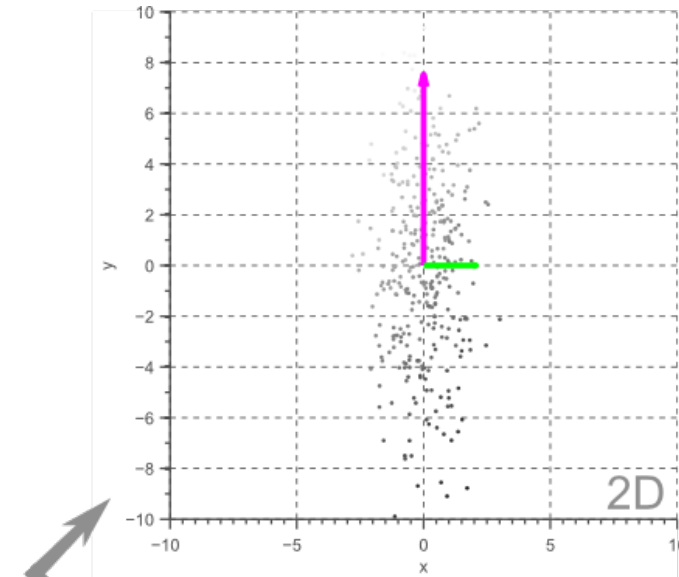
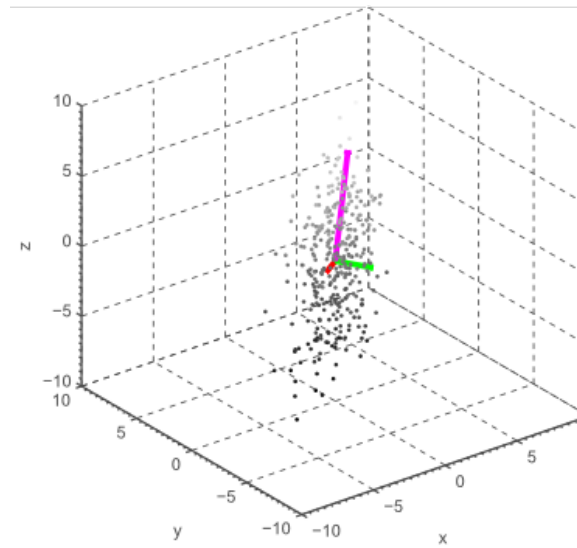


Faz-se o *standardization* dos dados:

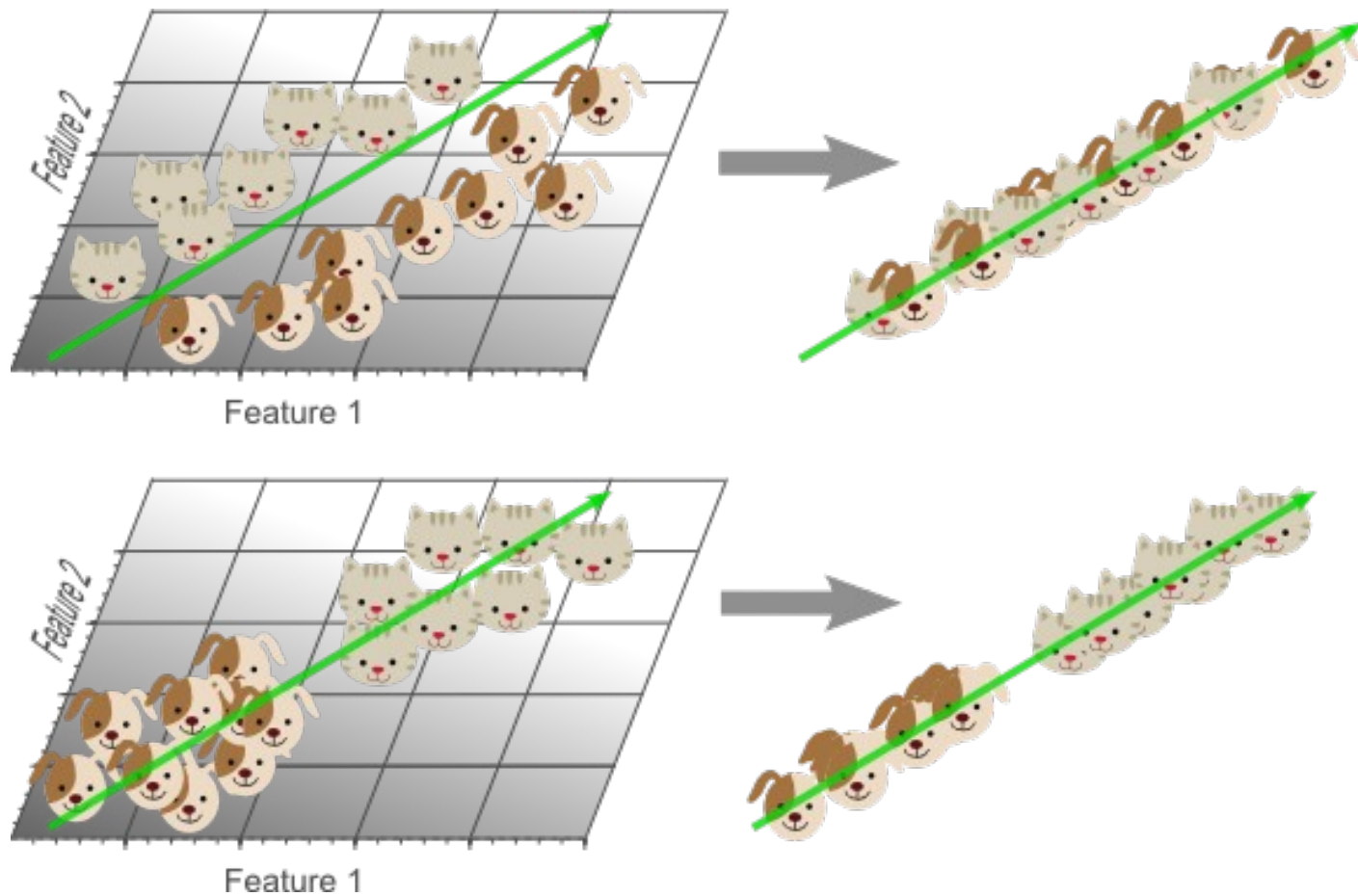


# PCA

Então, reduzimos sua dimensionalidade:



# PCA – Problemas encontrados



# Algoritmo do k-NN

## PASSOS:

1. Um inteiro positivo  $k$  é selecionado, igualmente para todo exemplo novo.
2. Calcula-se a distância do exemplo novo ( $x_q$ ) com cada exemplo ( $x_j$ ) do conjunto de treino, em todas dimensões  $i$ .

$$L^p(x_j, x_q) = \left( \sum_i |x_{j,i} - x_{q,i}|^p \right)^{1/p}$$

$p=1$  ← distância de Manhattan  
 $p=2$  ← distância Euclidiana

3. Seleciona-se  $k$  padrões no conjunto de treino que são mais próximos a este novo exemplo.
4. Seleciona-se a classificação mais comum dessas entradas (votação).
5. O novo padrão recebe essa classificação

# Distância de Minkowski ou norma $L^p$

$$L^p(x_j, x_q) = \left( \sum_i |x_{j,i} - x_{q,i}|^p \right)^{1/p}$$

- $x_q$  ponto de consulta;
- $x_j$  ponto de exemplo;
- Valores de  $p$ :

$p=1$  → distância de Manhattan.

- Usada se estiver medindo propriedades diferentes, como idade, peso, etc.
- Se os atributos forem booleanos, o número de atributos em que dois pontos diferem é chamado de distância de Hamming.

$p=2$  → distância Euclidiana

- Usada se estiver medindo propriedades similares, como largura, altura e profundidade.