

# Análise de Componentes Principais

---

Pontifícia Universidade Católica de Campinas

Prof. Dr. Denis M. L. Martins

# *Disclaimer*

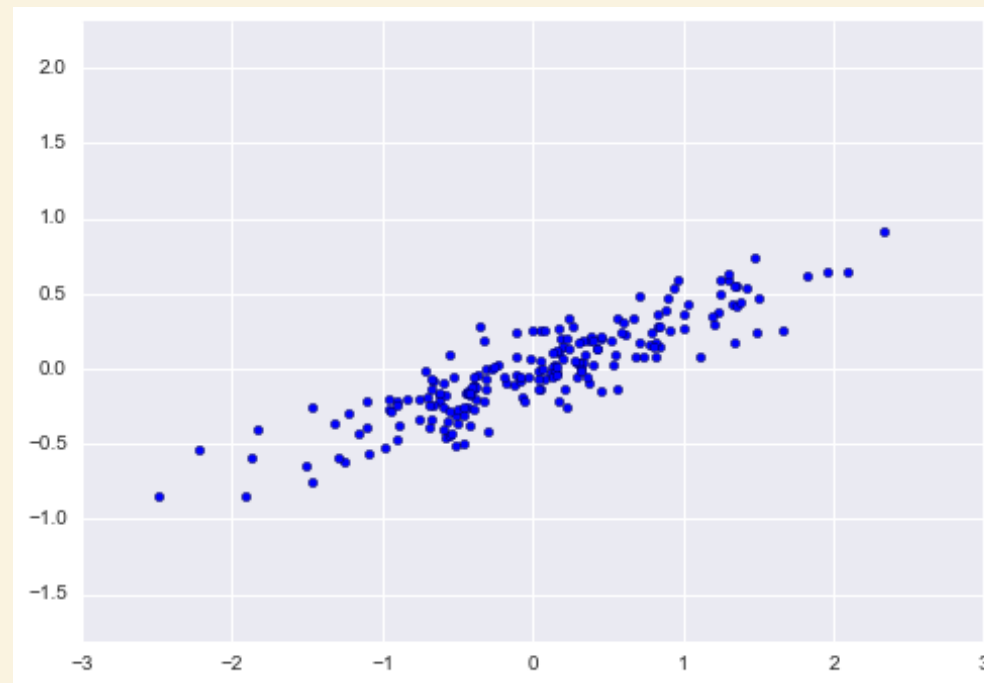
---

O material apresentado a seguir foi adaptado de [Jake VanderPlas](#).

# Análise Componentes Principais

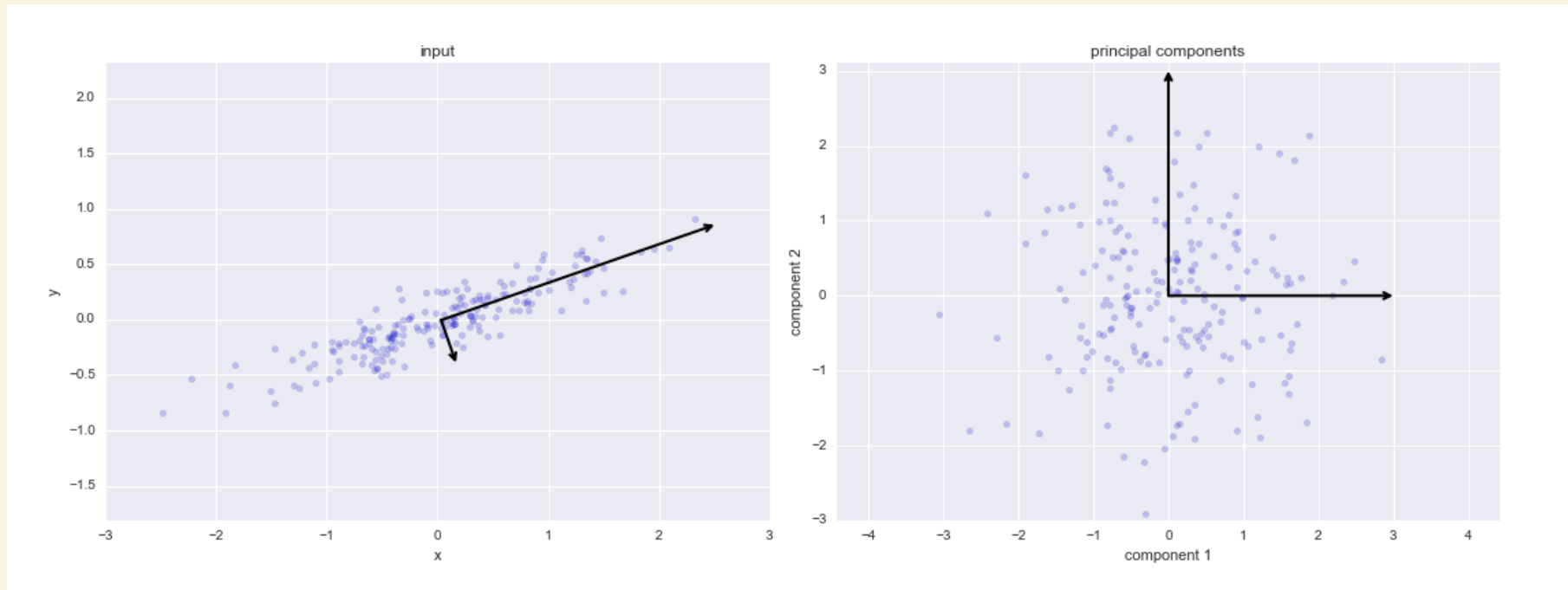
A Análise Componentes Principais (PCA) é um método não-supervisionado (utiliza dados não-rotulados) e rápido para redução de dimensionalidade em dados.

Considere o exemplo abaixo: relação linear aproximada entre as variáveis  $x$  e  $y$ .



# Análise Componentes Principais (cont.)

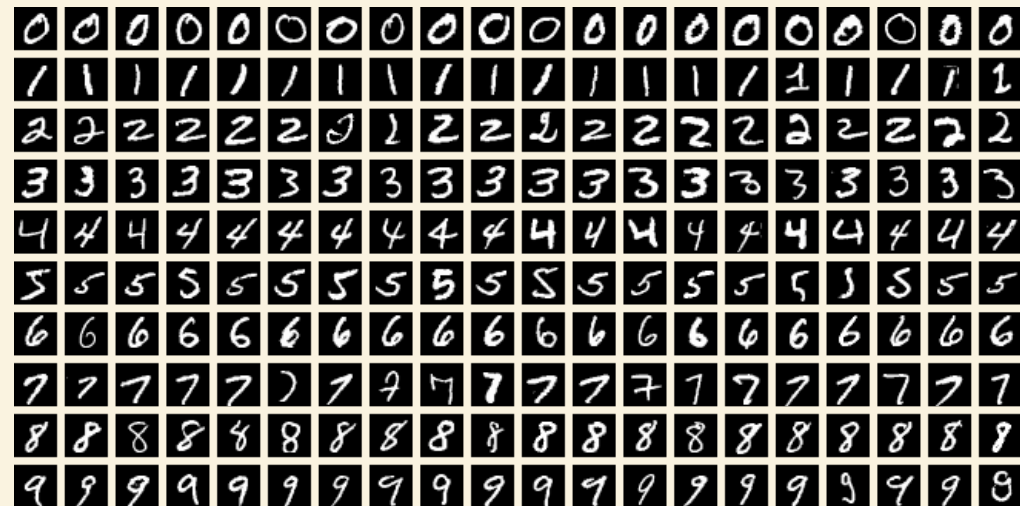
Na PCA, essa relação é quantificada através da identificação dos *eixos principais*, que são usados para descrever o conjunto de dados.



A projeção de cada ponto de dados nos eixos principais são os *componentes principais* dos dados.

# PCA para Visualização

Considere o MNIST dataset que contém imagens de 8x8 pixels referentes a dígitos escritos à mão.



PUC-Campinas | Prof. Dr. Denis Martins | [denis.mayr@puc-campinas.edu.br](mailto:denis.mayr@puc-campinas.edu.br)

# PCA para Visualização (cont.)

Podemos plotar os dois primeiros componentes principais de cada imagem para visualizar os dados:



# Quantos Componentes Escolher?

Estimar quantos componentes são necessários para descrever os dados: quantificar quanto da variância total de 64 dimensões está contida nos primeiros  $N$  componentes.

