

Latent Spaces in User Preference Modelling

André Uratsuka Manoel

Orientador: Gustavo Mirapalheta

Agenda

- ▶ Contexto
 - ▶ Preferência de Usuários
 - ▶ Espaços Latentes
 - ▶ Usos de Espaços Latentes
- ▶ Pergunta de Pesquisa
- ▶ Metodologia
 - ▶ MovieLens
 - ▶ Cálculo
- ▶ Resultados
- ▶ Conclusões

4 Desafios para a Compreensão de Preferências de Usuários

Informação Imperfeita

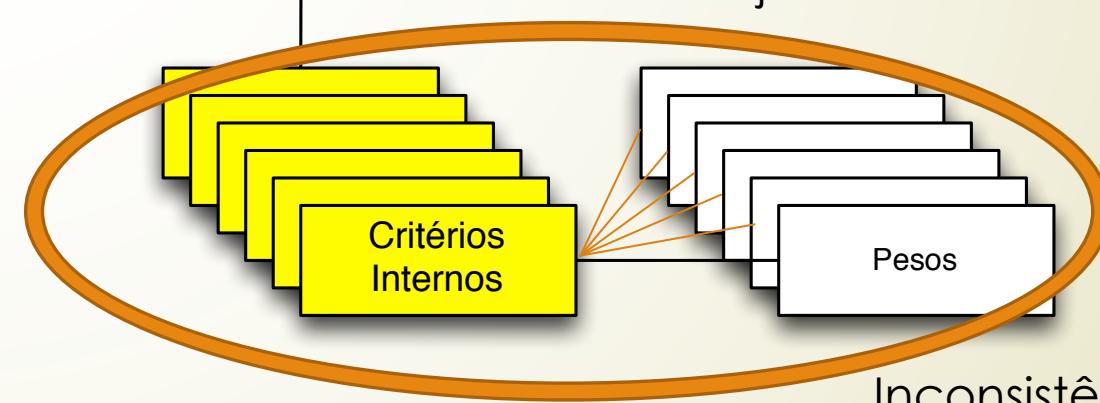
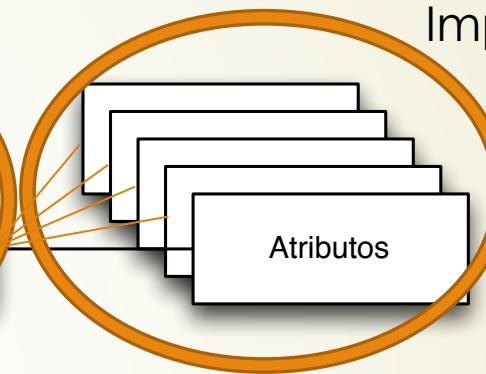
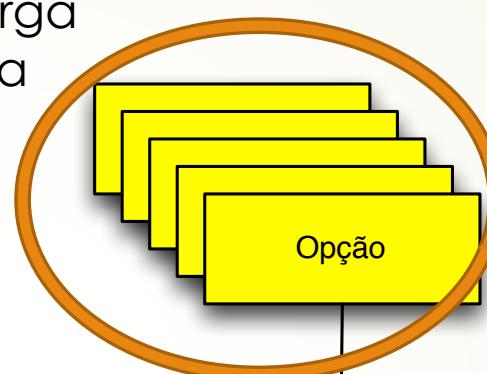
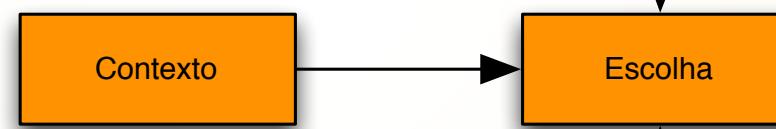
Sobrecarga Cognitiva

Subjetividade

Inconsistência

Sobrecarga Cognitiva

Informação Imperfeita



Subjetividade

Pesos

Inconsistência

Fonte: Elaborado pelo autor (2017)

Espaços Latentes

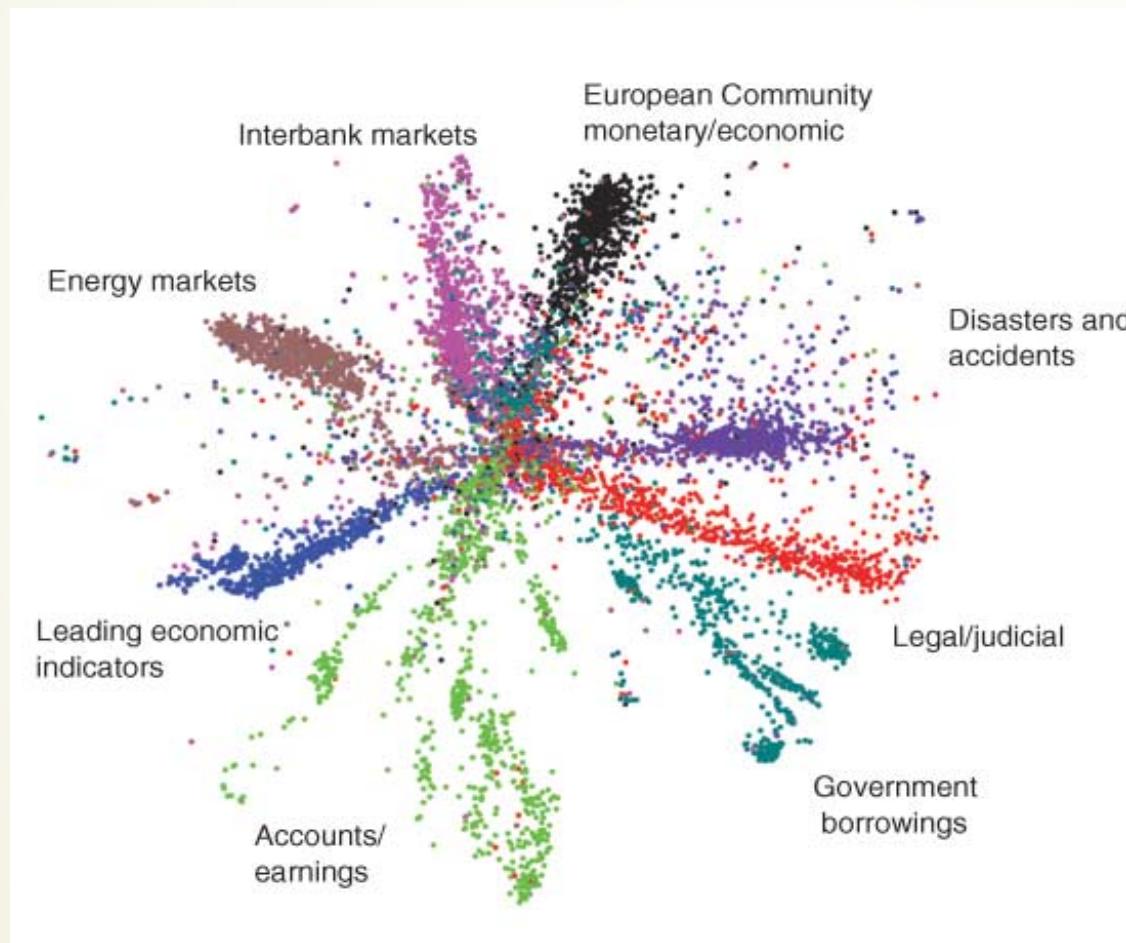
Espaços Vetoriais

- Cada componente corresponde a uma variável latente.
- Cada vetor representa um objeto e pode ser **incorporado (embedded)** em outros modelos

Construção

- Em ML: **Redes Neurais**
- Técnicas tradicionais: **PCA** e **Análise Fatorial**

Vetores Latentes em Machine Learning: Information Retrieval

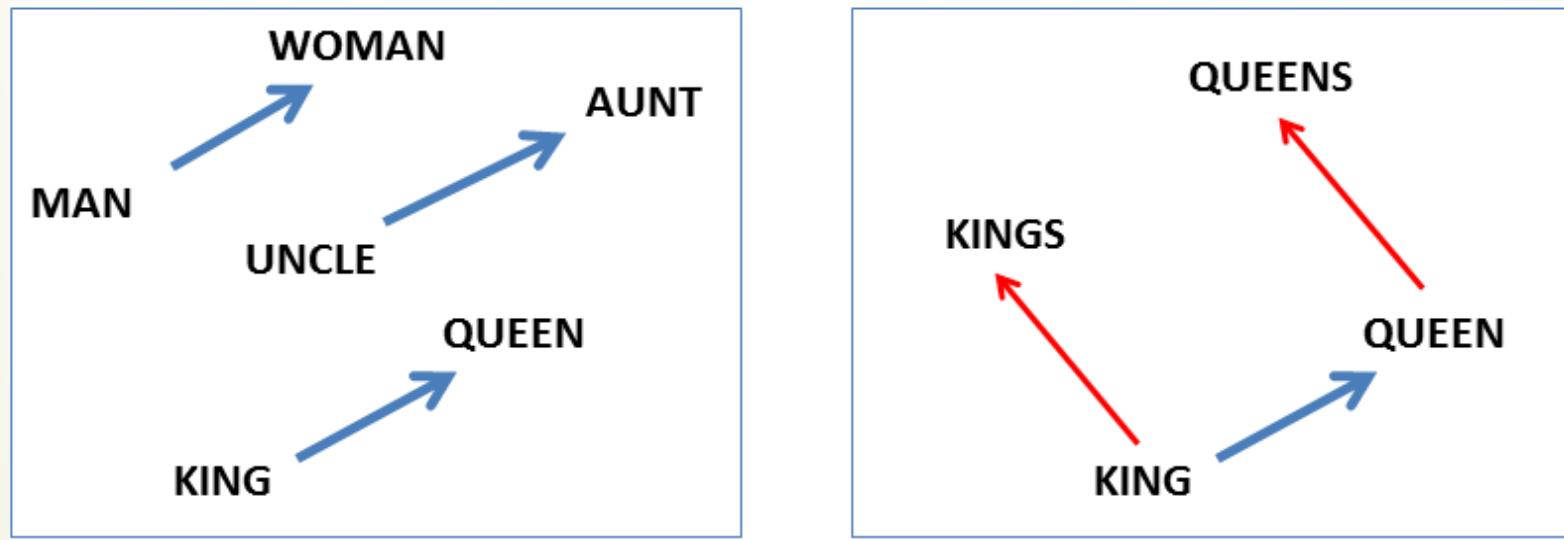


- Hinton; Salakhutdinov (2006): documentos são convertidos em vetores por uma Rede Neural RBM (**Máquina de Boltzmann Restrita**)
- Vetores tem estrutura que reflete temas

Fonte: Hinton; Salakhutdinov (2006, p. 506)

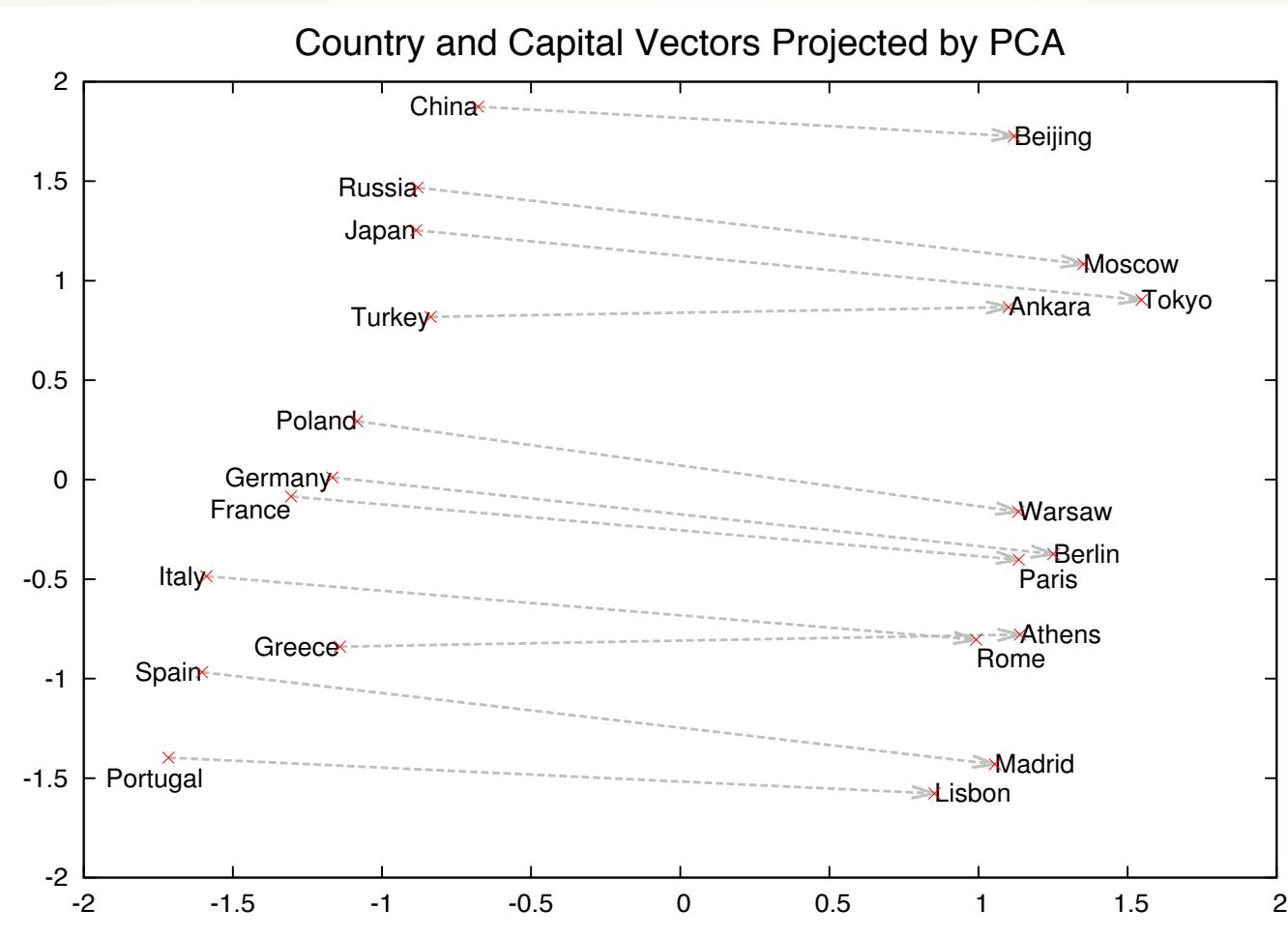
Word Embeddings

- Estrutura Emergente em Vetores de Palavras (Mikolov, 2013a)
- Mikolov et al. rede neural que modelava probabilidades de palavras aparecerem juntas, treinando palavras.
- "Gênero" e "plural" são direções no espaço latente gerado



Fonte: Mikolov et al. (2013a, p. 749)

Codificação semântica



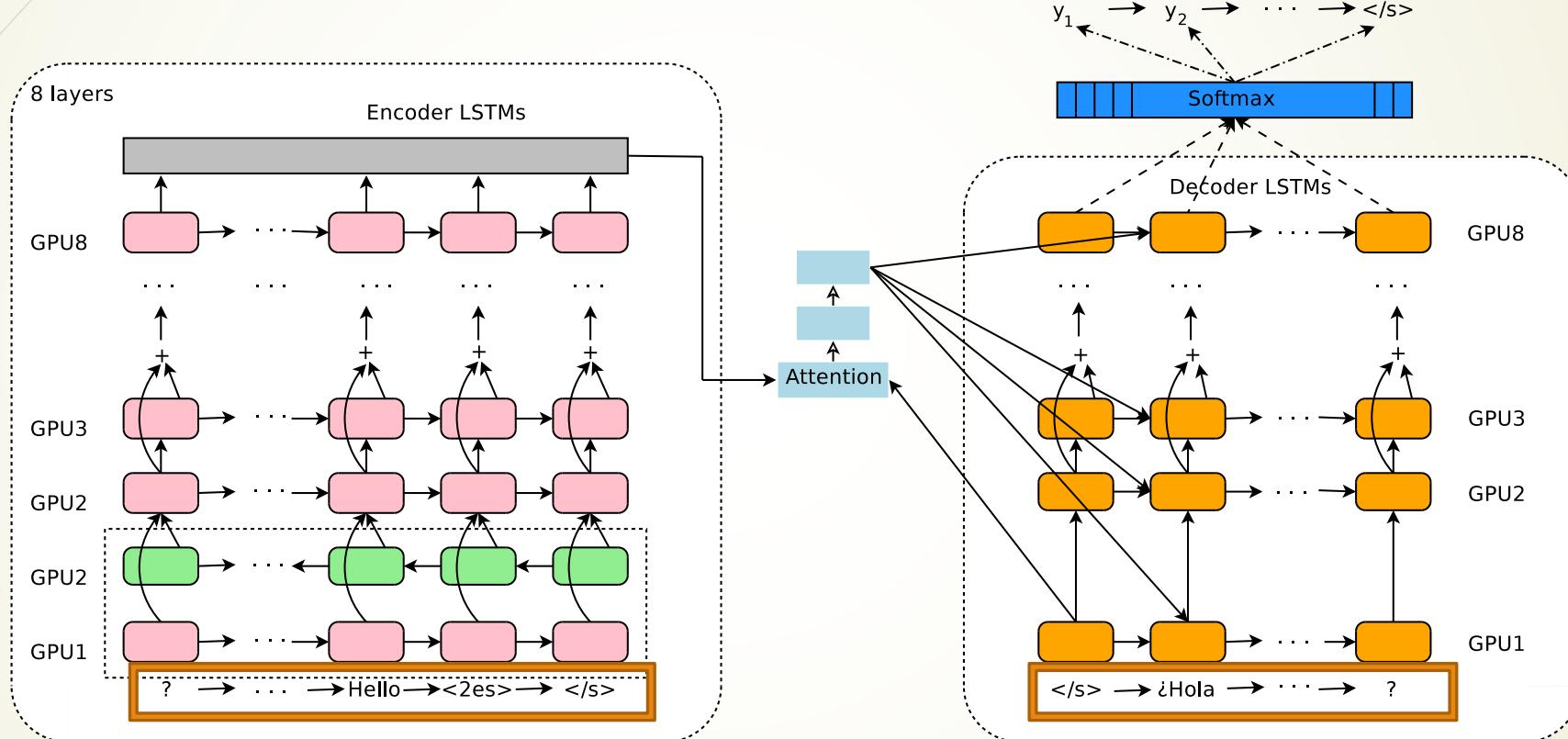
Fonte: Mikolov et al.
(2013c, p.4)

Publicação de word2vec

- Mikolov et al. (2013b):
 - Métodos para geração de vetores (bow, skipgram)
 - Publicação de Vetores pré-treinados **Word2Vec**

Fonte: Mikolov et al.
(2013c, p.4)

Vetores de Embedding em Tradução



Modelo de Tradução do Google Tradutor
Fonte: Johnson et al. (2017, p.4)

Pergunta de Pesquisa

► "É possível utilizar Espaços Latentes e outras técnicas recentes de Machine Learning para entender Preferências de Usuários?"

Metodologia

Objetivo

Criar vetores de usuários e de filmes da base **MovieLens**

Modelo

Fatoração de Matriz: $r_{ij} = u_i \cdot m_j + b_i^u + b_j^m$

Implementado em python com tensorflow, pandas e numpy (+Keras)

Minimizar

Erro Quadrático Médio na base de treinamento com **Batch Gradient Descent**

Validação

Base de treinamento, validação e testes

Monitorando

Erro Quadrático Médio na base de validação

Evitando Overfitting por

Regularização, Early Stopping

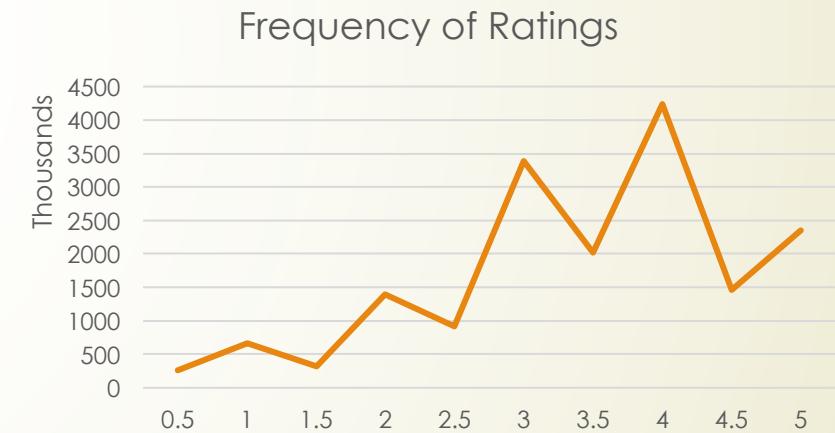
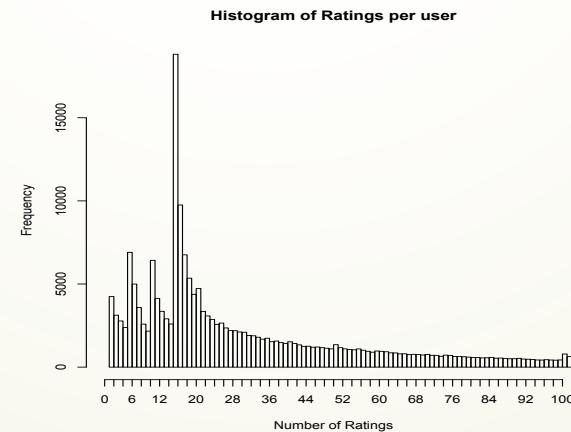
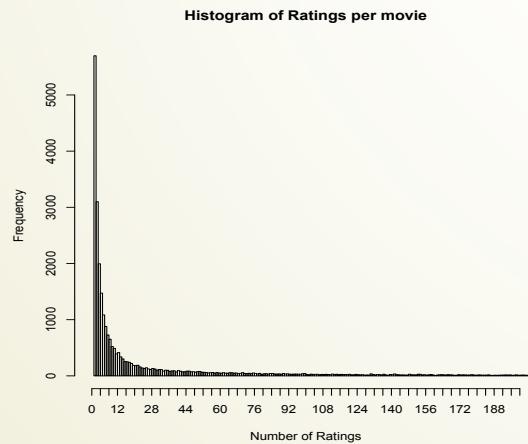
Análise

filmes similares por **distância cosseno**

Base MovieLens

- Base **MovieLens** de filmes (Maxwell; Konstan, 2015)
- Grupo GoupLens, Universidade de Minesotta

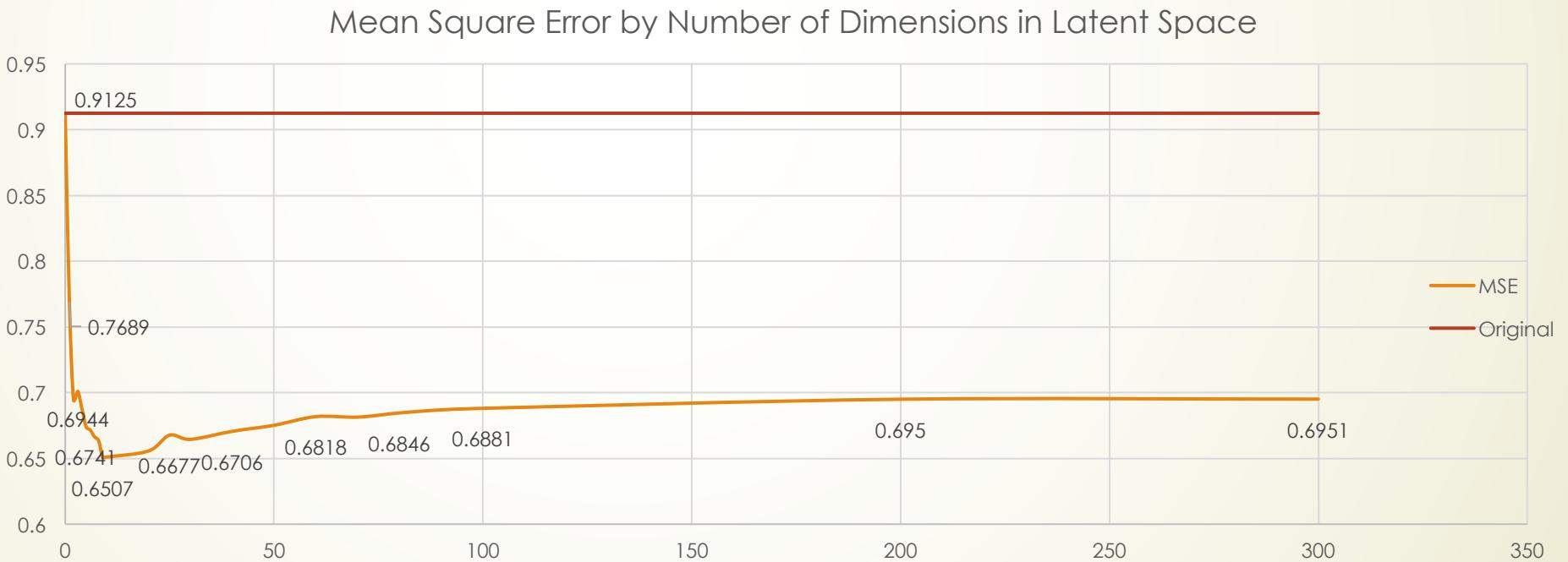
- **247.753** usuários x **33.670** filmes =
 - **8,3 bilhões** de avaliações possíveis
- **22.484.377** avaliações (**0,3%**)



Fonte: Elaborado pelo autor (2017)

Resultado

- Erro quadrático médio, utilizando apenas a média de notas é **0,9125**
- Devido a quantização, EQM tem piso de ~**0,3**
- **r²** de até 0,29



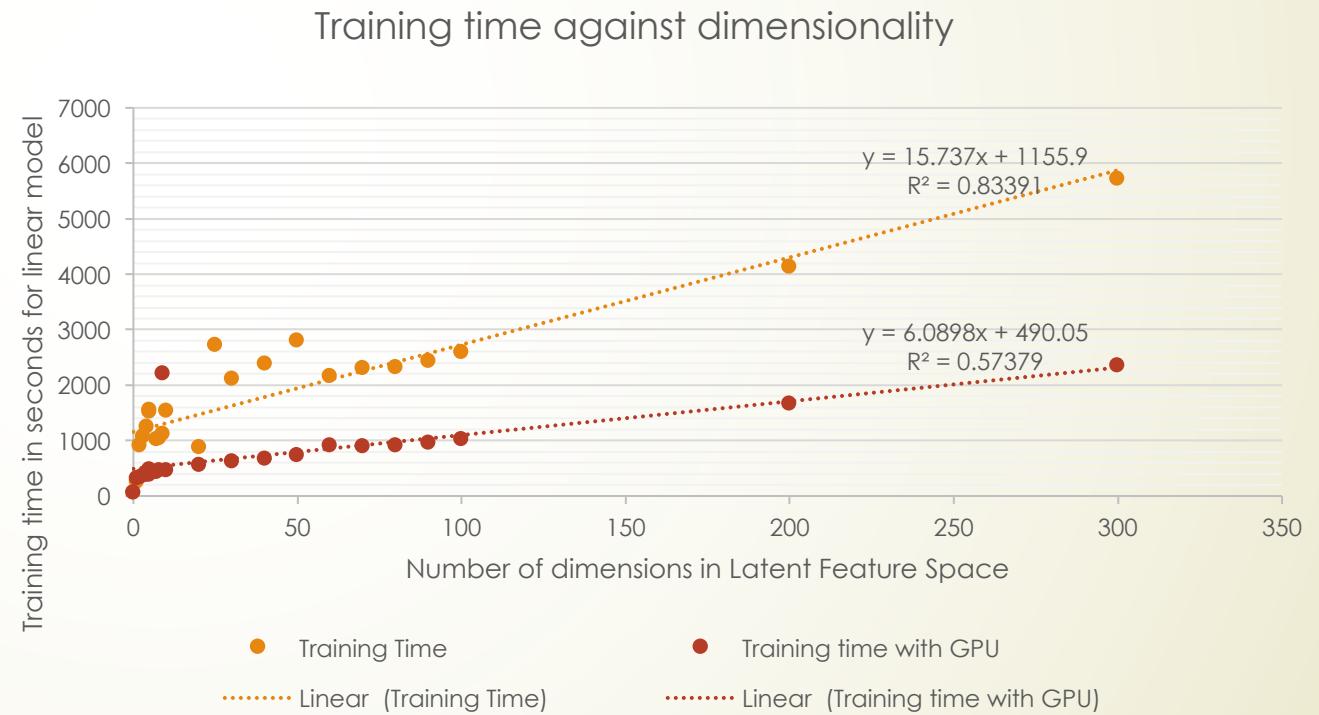
Fonte: Elaborado pelo autor (2017)

Performance é linear em k

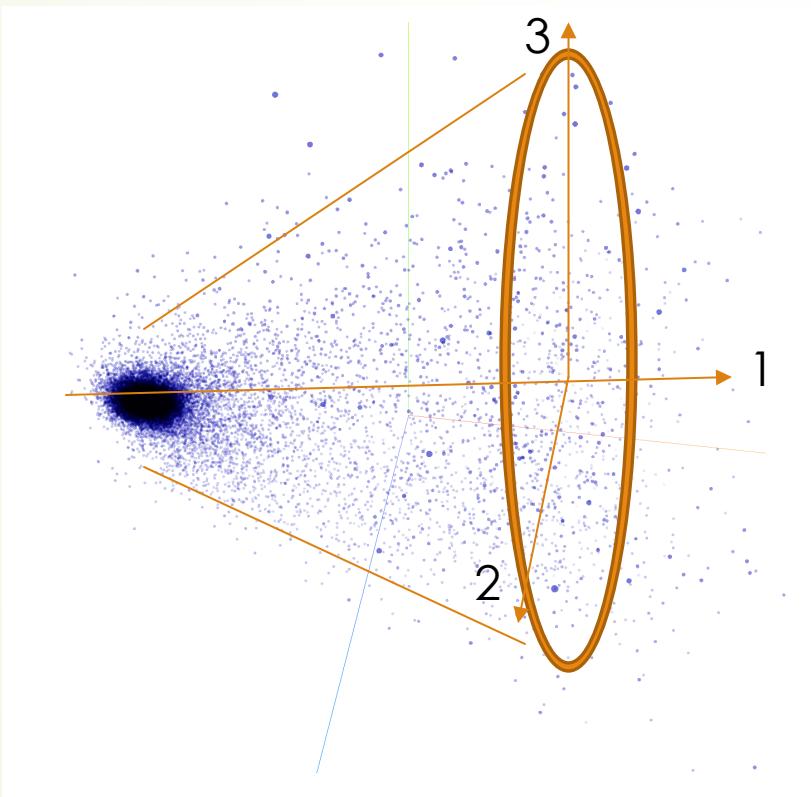
► Processamento em CPU (8 cores) e GPU (K80)

**Tempo de treinamento
aumenta linearmente com
número de dimensões**

Fonte: Elaborado pelo autor(2017)

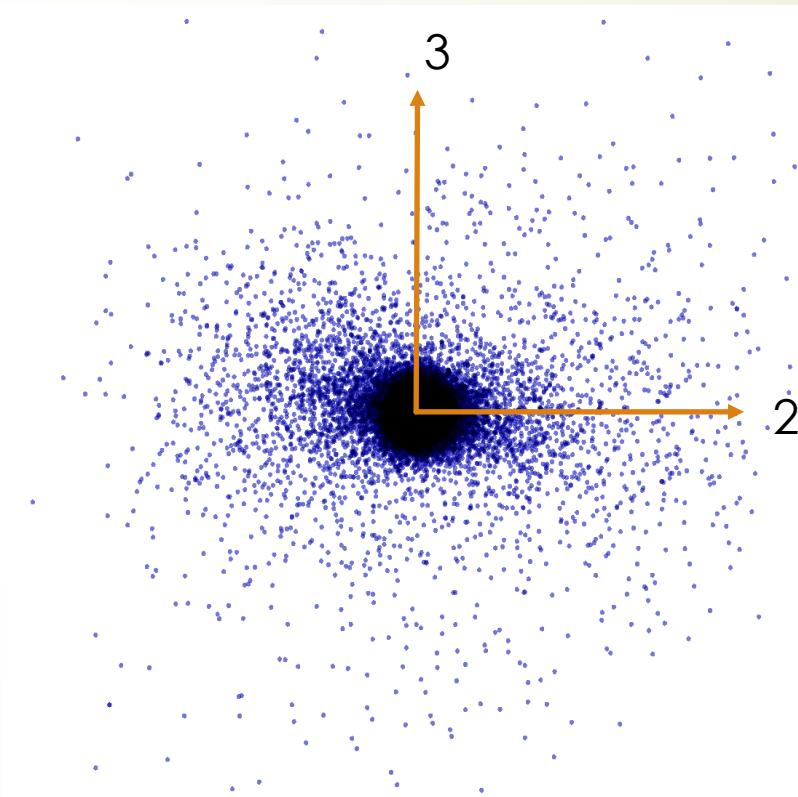


Estrutura tridimensional do espaço de filmes: 3 primeiros componentes PCA



Três primeiros componentes.

Fonte: elaborado pelo autor (2017)

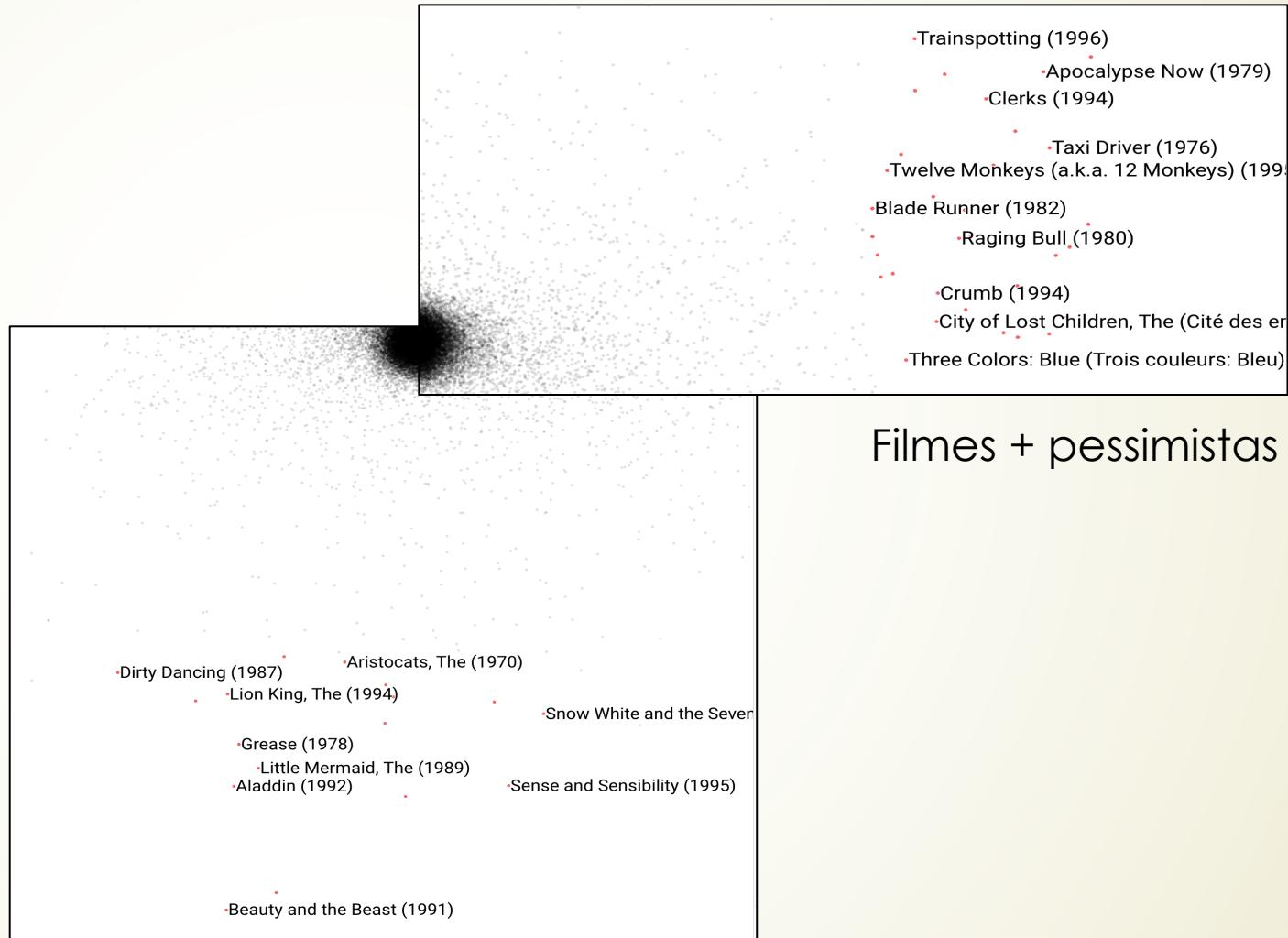


Segundo x terceiro componente.

Fonte: Elaborado pelo autor (2017)

Em 300 dimensões (projeção em 2 D)

Filmes + leves



Fonte: Elaborado pelo autor (2017)

Filmes mais Similares (3 dimensões)

Guerra nas Estrelas Episódio IV

Ferris Bueller's Day Off

Mulholland Falls

Planes, Trains & Automobiles

Garden State

Defending Your Life

Joe's Apartment

Sideways

Short Circuit 2

Fonte: Elaborado pelo autor (2017)



Mais dimensões melhoram resultados: filmes mais similares a

Guerra nas Estrelas: Episódio IV

3 dimensions	10 dimensions	300 dimensions
Ferris Bueller's Day Off	Star Wars Episode V	Star Wars Episode V
Mulholland Falls	Star Wars Episode VI	Star Wars Episode VI
Planes, Trains & Automobiles	Lord of The Rings: The Fellowship of the Ring	Raiders of the Lost Ark
Garden State	Raiders of the Lost Ark	Indiana Jones and the Last Crusade
Defending Your Life	Exotica	Star Wars: Episode III
Joe's Apartment	Lord of the Rings: The Two Towers	Star Wars: Episode VII
Sideways	Lord of the Rings: The Return of the King	Star Wars: Episode II
Short Circuit 2	Battlestar Galactica	Star Wars: Episode I

Fonte: Elaborado pelo Autor (2017)

Filmes Similares Compartilham Temas

Harry Potter and the Sorcerer Stone	Death Proof	Mississippi Burning	Footloose
Harry Potter and the Chamber of Secrets	Grindhouse	Boyz N The Hood	Flashdance
Harry Potter and the Goblet of Fire	Planet Terror	The Last King of Scotland	An Officer and a Gentleman
Harry Potter and the Prisoner of Azkaban	The Wrestler	All the President's Men	Free Willy 2
Harry Potter and the Order of the Phoenix	Inglourious Bastards	Kramer vs Kramer	Three Men and a Baby
Harry Potter and the Half-Blood Prince	Kill Bill, Vol. 2	Good Morning, Vietnam	The Karate Kid
Harry Potter and the Deathly Hallows: Part 1	In Bruges	A Bronx Tale	Steel Magnolias
Harry Potter and the Deathly Hallows: Part 2	Amores Perros	My Left Foot	Father of the Bride
The Hunger Games: Catching Fire	Django Unchained	Ray	Anastasia
Chronicles of Narnia: The Lion, the Witch, and the Wardrobe	Kill Bill, Vol. 1	Awakenings	
Tema: Harry Potter / Fantasy	Tema: Black Comedy, Violence, Tarantino	Tema: Luta, esperança, opressão	Tema: Libertaçāo / Romance / Emoção

Fonte:
Elaborado pelo
autor (2017)

Conclusões

- ▶ Fatoração de Matrizes é um mecanismo eficiente e eficaz para gerar espaços latentes (mas não o único)
- ▶ Com K grande o suficiente há aprendizado de características sutis de filmes e usuários
- ▶ Os vetores podem ser usadas para alimentar modelos mais complexos e fazer previsões

Usos da técnica

- ▶ Entendimento de produtos e **preferências de consumidor**
- ▶ **Sistemas de Recomendação**
- ▶ **Estimar variáveis em populações**, levando em conta características específicas ao invés de médias
- ▶ **Avaliação de ativos**: vetores de empresas e de situação

Limitações

- ▶ Vetores são **estimativa pontual** (regressão **bayesiana** ou **processo gaussiano** poderiam ser mais apropriados em alguns situações)
- ▶ Não existe um **entendimento** do que significa cada componente do vetor
- ▶ Fatoração de Matrizes não explora a **não-linearidade** do espaço
- ▶ Exige um **volume muito grande de dados** (embora significado possa ser extraído de dados que não eram considerados úteis. Por exemplo: cotações)

Pesquisa Futura

- ▶ Adicionar **outros dados** sobre filmes e usuários
- ▶ **Transportar** gosto para filmes para outros produtos
- ▶ Estimar **grau de certeza** em vetores
- ▶ Tratar **variação temporal**
- ▶ Estudar **significado de componentes**
- ▶ Outras aplicações: **finanças, análise de mercados**



Obrigado

Referências

- ▶ Adamavicius, G., Tuzhilin, A. (2005). Personalization Technologies: A Process-oriented Perspective. *Communications of the ACM*, 48(10), 83-90.
- ▶ Hinton, G. E. & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science Magazine* (313), p. 504-507. DOI: 10.1126/science.1127647. Retrieved Nov 12, 20017 from <https://pdfs.semanticscholar.org/7d76/b71b700846901ac4ac119403aa737a285e36.pdf>
- ▶ Hoong, K. (2016). [Data Science, IoT, Big Data Analytics, Machine Learning](https://kuanhoong.wordpress.com/2016/02/01/r-and-deep-learning-cnn-for-handwritten-digits-recognition/). [Blog Post]. Disponível em <https://kuanhoong.wordpress.com/2016/02/01/r-and-deep-learning-cnn-for-handwritten-digits-recognition/>
- ▶ Jolliffe, I. T. (2002). Principal Component Analysis. New York: Springer
- ▶ Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. <https://doi.org/10.1109/MC.2009.263>
- ▶ Mikolov, T., Yih, W., & Zweig, G. (2013a). Linguistic regularities in continuous space word representations. *Proceedings of NAACL-HLT*, (June), 746–751.
- ▶ Mikolov, T., Corrado G., Chen, K. Dean, J. (2013b, September 7). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.13781v3 [cs.CL]*. Retrieved November 27, 2017.
- ▶ Mikolov, T., Sutskever, I. Chen, K., Corrado, G. & Dean J. (2013c, October 1). Distributed Representation of Words and Phrases and Their Compositionality. [arXiv:1310.4546](https://arxiv.org/abs/1310.4546) [cs.CL]. Retrieved November 14th, 2017.



Apêndices

Modelos básicos de recomendação

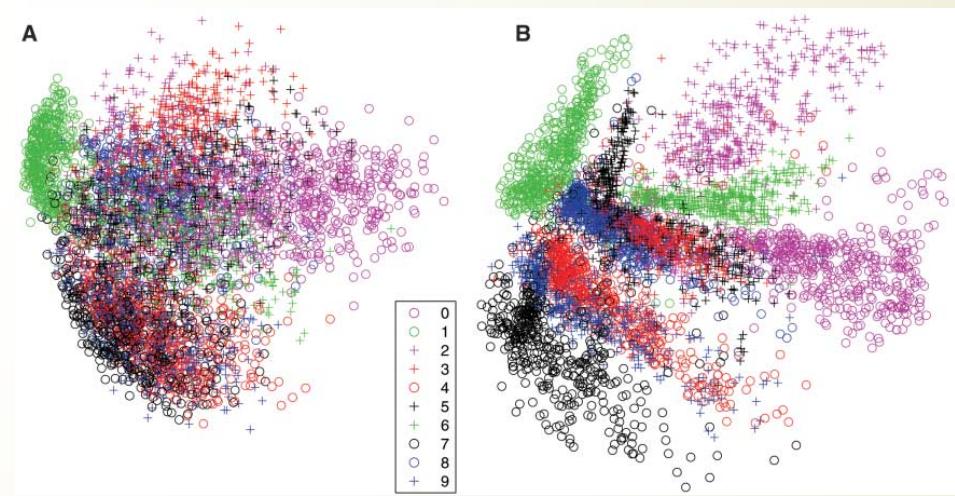
- ▶ Baseados em conteúdo
- ▶ Baseados em usuários (filtros colaborativos)
- ▶ Híbridos

pandora Google

amazon Spotify

Vetores de Embedding em Machine Learning: reconhecimento de texto

Fonte: Hoong, K. (2016).



(A) Usando PCA simples; (B) Usando Autoencoder.
Fonte: Hinton; Salakhutdinov (2006, p. 506)

Origem de Modelos de Variáveis Latentes

- ▶ “Latent variable models and factor analysis are **among the oldest multivariate methods**, but their origin and development lie almost entirely **outside of mainstream statistics**. For this reason topics such as latent class analysis and factor analysis had separate origins and have remained almost totally isolated from one another. This means that they have been regarded as separate fields with virtually no cross-fertilisation.”

(Bartholomew et al., 2001, p. 12)

PCA e Análise Fatorial

- ▶ Uma das técnicas mais comuns de análise de populações
- ▶ PCA: Criado por **Pierson** (1901), **Hotelling** (1933)
- ▶ Análise Fatorial: múltiplos autores, evoluindo independentemente ao longo dos anos
- ▶ Análise Fatorial: $\mathbf{X} = \Lambda \mathbf{f} + \mathbf{e}$ (Jolliffe, p. 151)
- ▶ PCA: transformações ortogonais nas variáveis. Usualmente se usa SVD
- ▶ Saída do modelo: uma base e os componentes para cada elemento sendo estudado, que pode ser interpretado como um vetor também

PCA via SVD

$$R = U\Sigma V^T$$

Dimensions indicated by arrows:

- $n \times m$ arrow points to U
- $n \times n$ arrow points to Σ
- $n \times m$ diagonal arrow points to V^T
- $m \times m$ arrow points to V^T

Text below:

- Em PCA, U , Σ e V^T são únicos

Fatoração de Matriz (4 formas equivalentes)

$$R_{ij} = \sum_{l=1}^k u_{il} m_{lj}$$
$$r_{ij} = \begin{bmatrix} u_{i1}, u_{i2}, u_{i3}, u_{i4}, u_{i5} \end{bmatrix} \times \begin{bmatrix} m_{1j} \\ m_{2j} \\ m_{3j} \\ m_{4j} \\ m_{5j} \end{bmatrix}$$
$$r_{ij} = u_i \cdot m_j$$

$$R = UM$$

Fatoração de Matriz com Viés

$$R_{ij} = \left[u_{i1}, u_{i2}, u_{i3}, u_{i4}, u_{i5}, 1, b_i^u \right] \begin{Bmatrix} m_{1j} \\ m_{2j} \\ m_{3j} \\ m_{4j} \\ m_{5j} \\ b_j^m \\ 1 \end{Bmatrix} \quad (k = 5)$$

$$r_{ij} = u_i \cdot m_j + b_i^u + b_j^m$$

(KOREN; BELL; VOLINSKY, 2009)

Similaridade entre vetores v e w

$$\text{Similaridade}(v, w) = \frac{v \cdot w}{\|v\|_2 \|w\|_2}$$

$$\text{cosine dist}(v, w) = 1 - \text{similaridade}(v, w)$$

Regularização

$$\begin{aligned} Perda &= BatchSqError + \lambda reg \\ reg &= \frac{\|u^T \cdot u\|^2}{k \cdot b} + \frac{\|m^T \cdot m\|^2}{k \cdot b} \end{aligned}$$

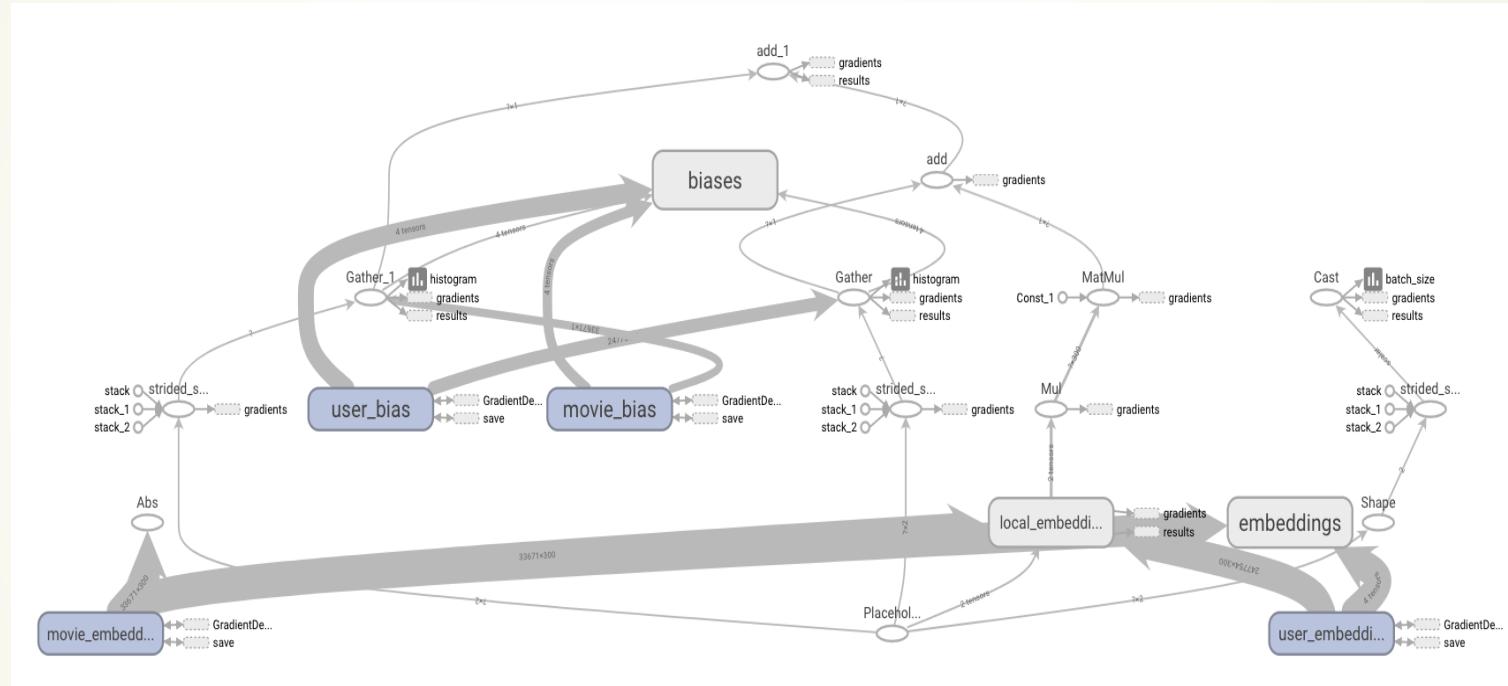
- k – número de dimensões
- b – tamanho do batch

Notas dadas por usuários formam Matriz esparsa

userId	movielid	rating
123432	13562	4.0
127322	23343	4.0
234	1	5.0
43234	1234	3.5
12123	21446	3.0



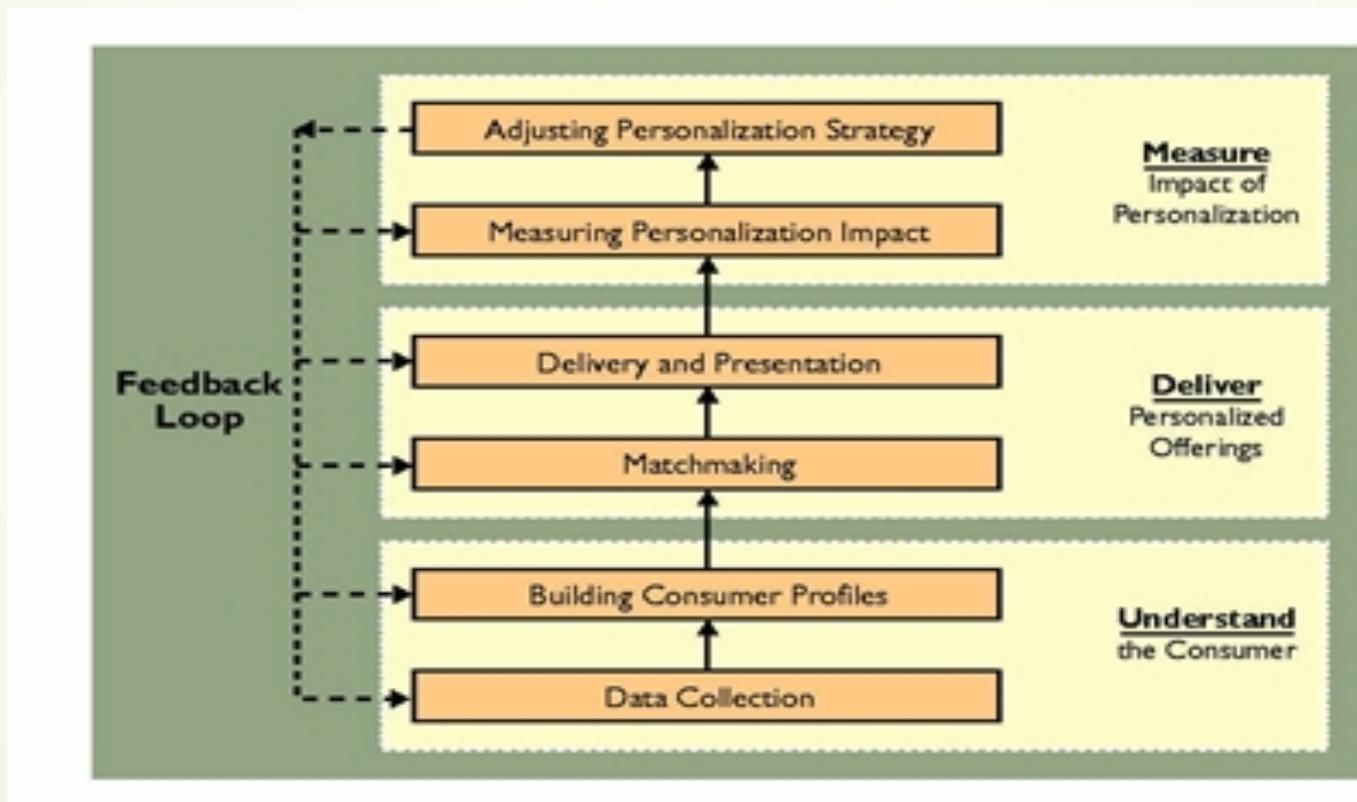
Modelo Tensorflow



Fonte: Elaborado pelo Autor (2017)

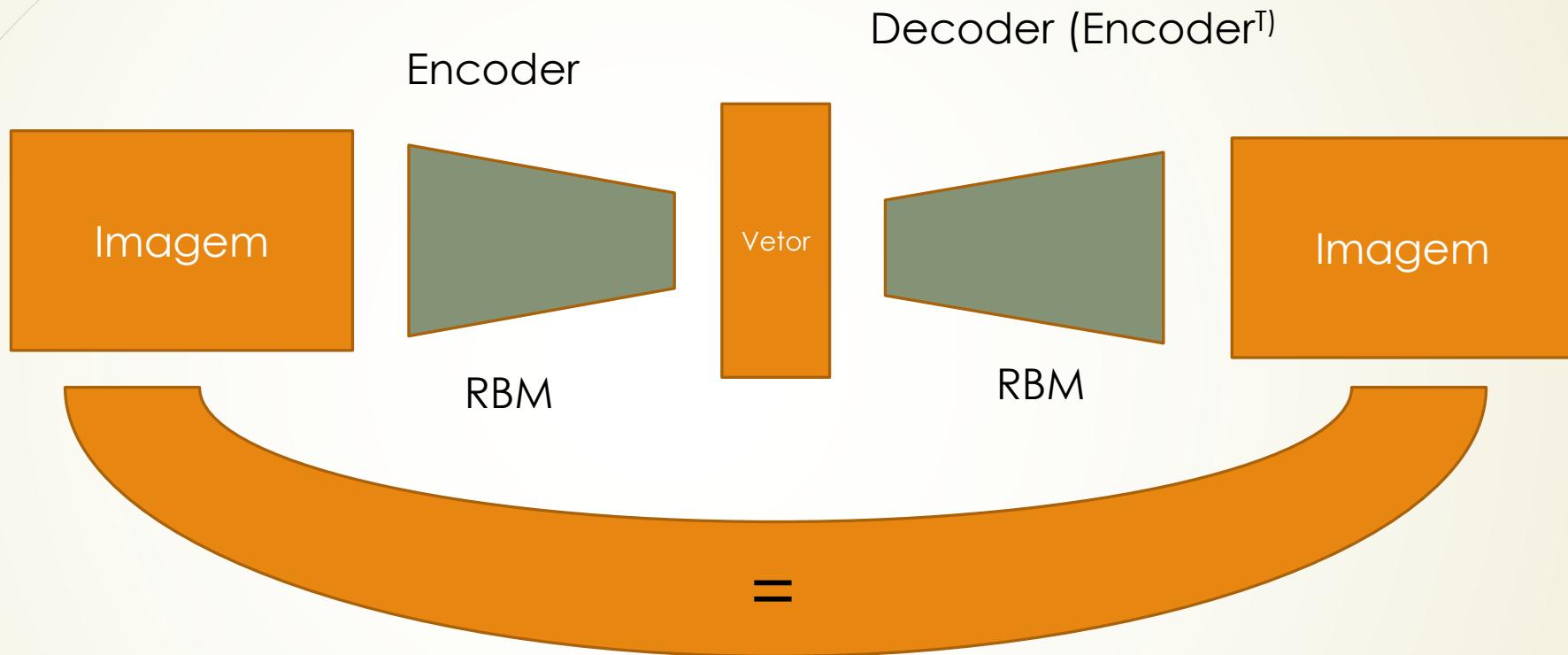
- Perda: Erro Quadrático Médio
- calculado somente na amostra
- python 3.7
- Tensorflow, numpy, pandas, scipy

Sistemas de Recomendação



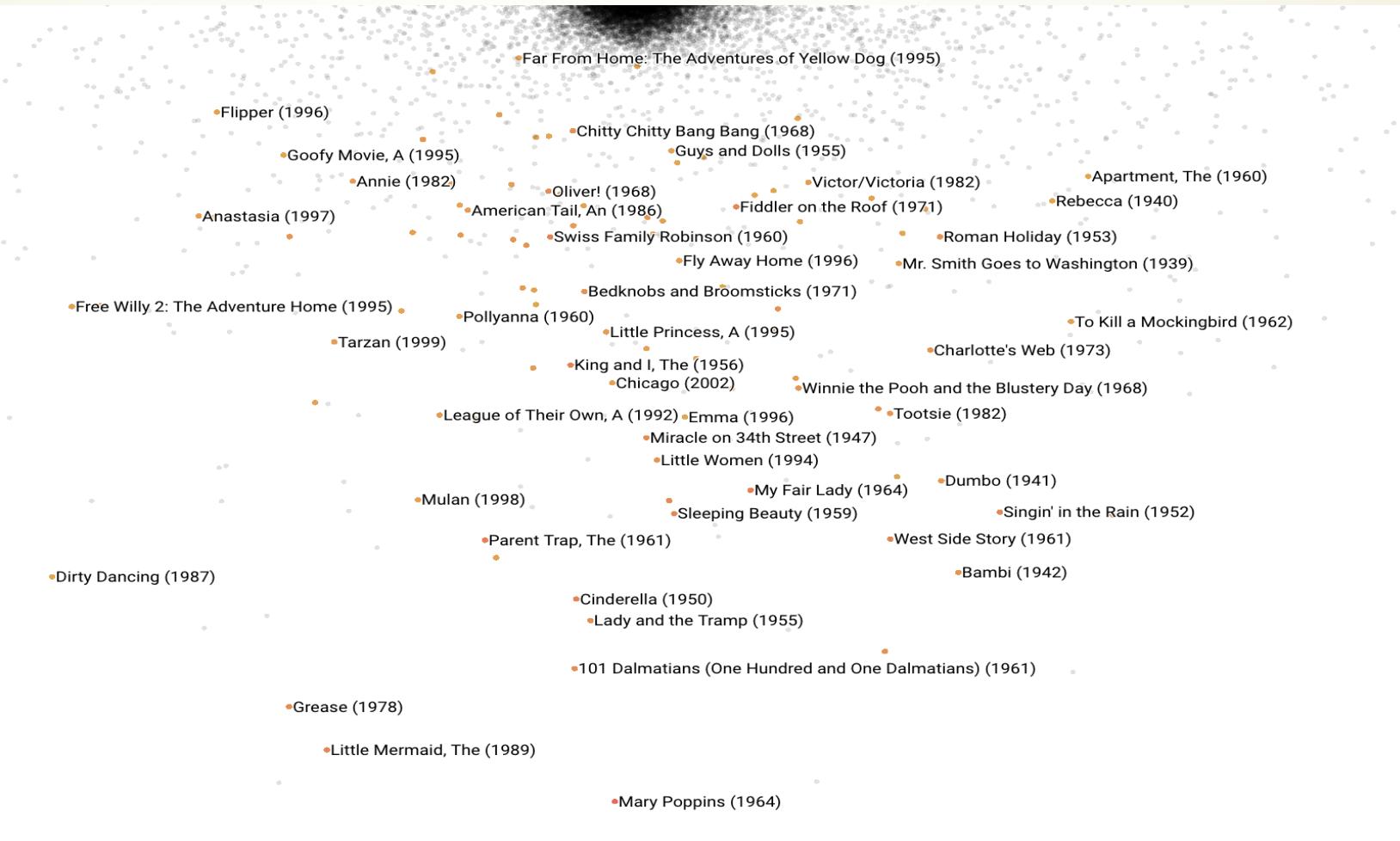
Fonte: Adamavicius; Tuzhlin (2005, p. 87)

Estrutura Autoencoder



Encoder é uma série de camadas contendo multiplicação de matrizes por pesos e aplicação de função de "ativação"

Vizinhança da **Noviça Rebelde** (300-d)

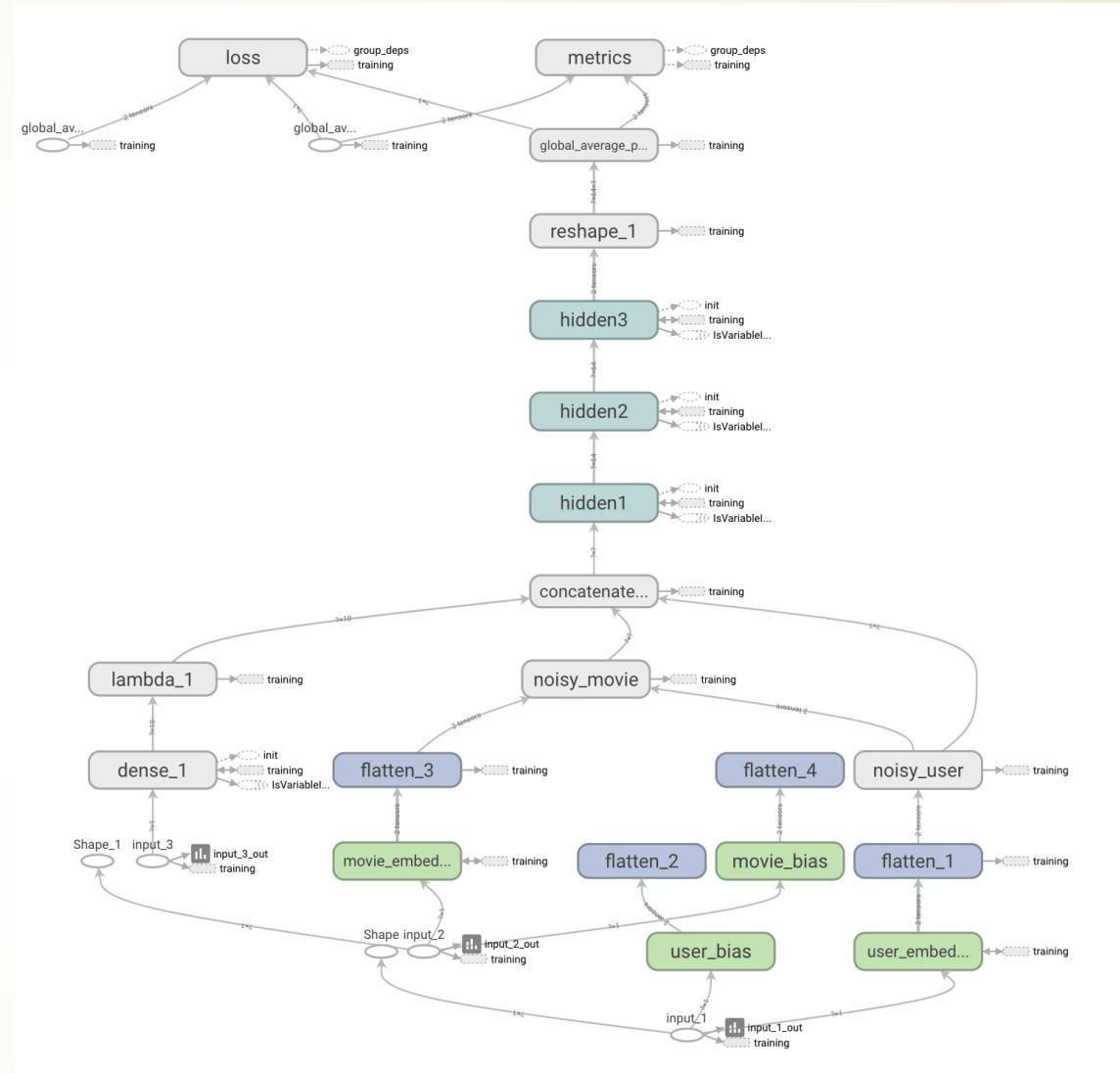


Fonte: Elaborado pelo Autor (2017)

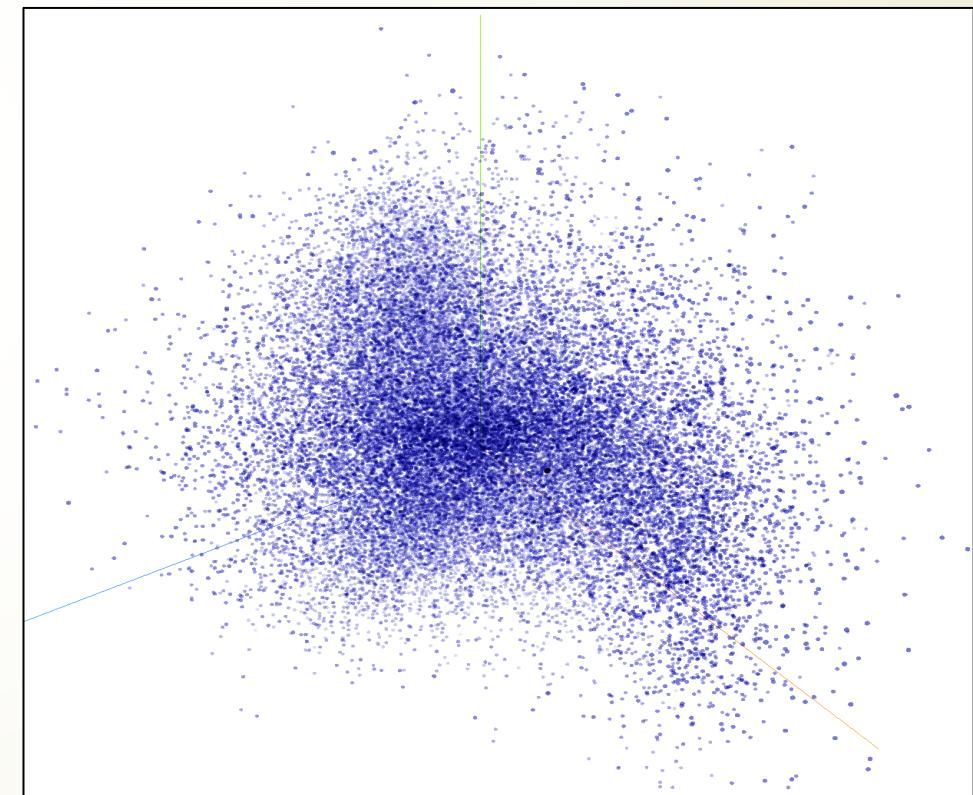
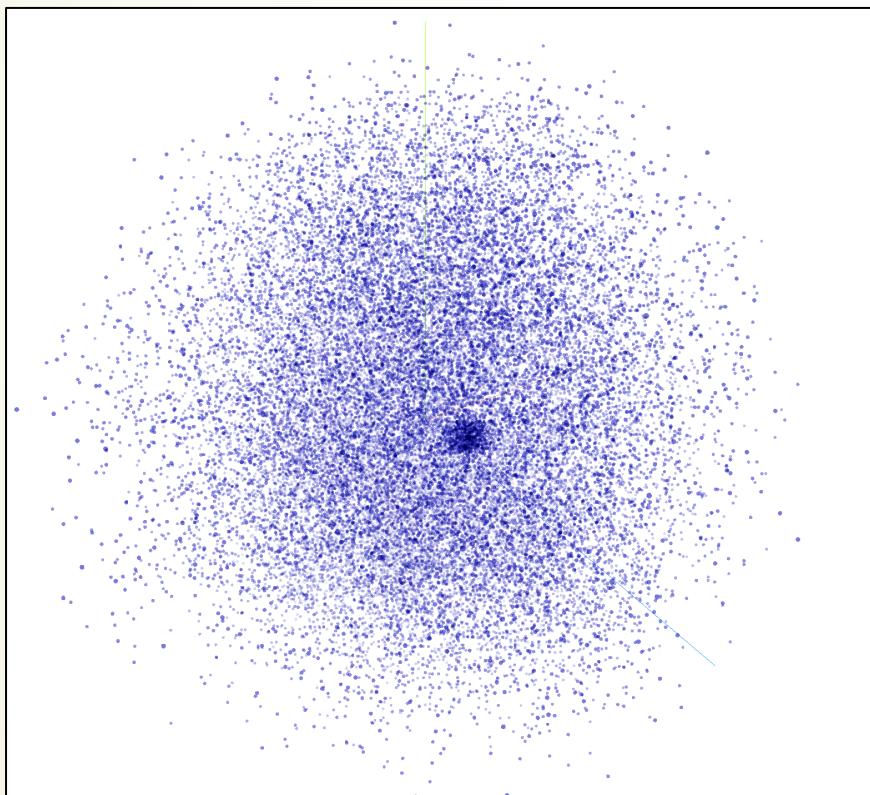
Modelo de Rede Neural

- O modelo de fatoração não é o único possível.
- Um modelo de rede neural com 4 camadas tem resultados similares a fatoração de matrizes
- Ruído Gaussiano para eviter overfitting

Fonte: Elaborado pelo autor (2017)



Espaços Gerados com Modelo de Rede Neural



Fonte: Elaborado pelo Autor (2017)