

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe Tasaka de Melo
Helyane Bronoski Borges

Universidade Tecnológica Federal do Paraná (UTFPR)

20/09/2019

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

Introdução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Aprendizados de máquina são cada vez mais utilizados nos dias atuais

Introdução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Aprendizados de máquina são cada vez mais utilizados nos dias atuais
- Uma das vantagens é sua capacidade de lidar com grandes bases de dados

Introdução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Aprendizados de máquina são cada vez mais utilizados nos dias atuais
- Uma das vantagens é sua capacidade de lidar com grandes bases de dados
- Exemplos de bases: Banco de dados de clientes, imagens de alta resolução, **microarranjos**

Introdução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Aprendizados de máquina são cada vez mais utilizados nos dias atuais
- Uma das vantagens é sua capacidade de lidar com grandes bases de dados
- Exemplos de bases: Banco de dados de clientes, imagens de alta resolução, **microarranjos**

Microarranjo

Técnica de capturação do DNA, utilizada no estudo da expressão genética

Microarranjo

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Um dos usos é na captação de células cancerígenas

Microarranjo

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Um dos usos é na captação de células cancerígenas
- O resultado é uma matriz numérica onde cada característica é uma coluna e cada linha uma amostra

Microarranjo

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Um dos usos é na captação de células cancerígenas
- O resultado é uma matriz numérica onde cada característica é uma coluna e cada linha uma amostra

attribute1	attribute2	attribute3	attribute4	attribute5	attribute6	attribute7	attribute8	attribute9	attribute10	attribute11
-214	-153	-58	88	-295	-558	199	-176	252	206	-41
-139	-73	-1	283	-264	-400	-330	-168	101	74	19
-76	-49	-307	309	-376	-650	33	-367	206	-215	19
-135	-114	265	12	-419	-585	158	-253	49	31	363
-106	-125	-76	168	-230	-284	4	-122	70	252	155
-138	-85	215	71	-272	-558	67	-186	87	193	325
-72	-144	238	55	-399	-551	131	-179	126	-20	-115
-413	-260	7	-2	-541	-790	-275	-463	70	-169	361
5	-127	106	268	-210	-535	0	-174	24	506	284
-88	-105	42	219	-178	-246	328	-148	177	183	-143
-165	-155	-71	82	-163	-430	100	-109	56	350	204
-67	-93	84	25	-179	-323	-135	-127	-2	-66	208
-92	-119	-31	173	-233	-227	-49	-62	13	230	2
-113	-147	-118	243	-127	-398	-249	-228	-37	113	188
-107	-72	-126	149	-205	-284	-166	-185	1	-23	205
-117	-219	-50	257	-218	-402	228	-147	65	67	-101
-476	-213	-18	301	-403	-394	-42	-144	98	173	-133
-81	-150	-119	78	-152	-340	-36	-141	96	-55	-209
-44	-51	100	207	-146	-221	83	-198	34	-20	11

Figura: Exemplo do trecho de uma base de microarranjo.
A autoria do autor

Problemas das bases

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- As bases são extremamente grandes

Problemas das bases

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- As bases são extremamente grandes
- A quantidade de instâncias é grande o suficiente para que o processamento de uma futura análise pode ser comprometida

Problemas das bases

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- As bases são extremamente grandes
- A quantidade de instâncias é grande o suficiente para que o processamento de uma futura análise pode ser comprometida
- Bases de microarranjo ainda possuem a **maldição da dimensionalidade**

Problemas das bases

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- As bases são extremamente grandes
- A quantidade de instâncias é grande o suficiente para que o processamento de uma futura análise pode ser comprometida
- Bases de microarranjo ainda possuem a **maldição da dimensionalidade**

Maldição da dimensionalidade

Para um número de instâncias p , a quantidade m de amostras é extremamente pequena em relação à p (GÉRON, 2017)

Maldição da dimensionalidade

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Escassez de dados gera falta de diversidade no aprendizado da IA e, conseqüentemente, previsões erradas

Maldição da dimensionalidade

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Escassez de dados gera falta de diversidade no aprendizado da IA e, conseqüentemente, previsões erradas
- Muitas instâncias pode dificultam a IA de encontrar um padrão entre as amostras

Maldição da dimensionalidade

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Escassez de dados gera falta de diversidade no aprendizado da IA e, conseqüentemente, previsões erradas
- Muitas instâncias pode dificultam a IA de encontrar um padrão entre as amostras
- A redução da dimensionalidade facilitará um futuro processamento e melhorará o diferenciamento entre os grupos

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

Técnicas de redução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

**Técnicas de
redução**

Redes neurais
autocodificadoras

Objetivo

Análise

Referências

- Duas grandes abordagens:

Técnicas de redução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de
redução

Redes neurais
autocodificadoras

Objetivo

Análise

Referências

- Duas grandes abordagens:
 - 1 Seleção de atributos: Visa selecionar os atributos mais importantes sem modificar a base original

Técnicas de redução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Duas grandes abordagens:
 - 1 Seleção de atributos: Visa selecionar os atributos mais importantes sem modificar a base original
 - 2 Extração de atributos: Seleciona os melhores atributos, contudo ele altera a base de dados original

Técnicas de redução

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Duas grandes abordagens:
 - 1 Seleção de atributos: Visa selecionar os atributos mais importantes sem modificar a base original
 - 2 Extração de atributos: Seleciona os melhores atributos, contudo ele altera a base de dados original
- Exemplos de métodos de redução: PCA, Fuzzy-Rough, entre outras.

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

O que é uma rede neural?

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Estrutura amplamente utilizada na área de deep learning

O que é uma rede neural?

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Estrutura amplamente utilizada na área de deep learning
- Procura "imitar" os neurônios humanos, ativando ligações importantes e desativando outros

O que é uma rede neural?

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Estrutura amplamente utilizada na área de deep learning
- Procura "imitar" os neurônios humanos, ativando ligações importantes e desativando outros
- Existem vários tipos: simples, convolucionais, adversariais, autocodificadoras.

Redes neurais autocodificadoras

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Busca, a partir de uma entrada, reduzi-la (codificação) e, depois, expandi-la novamente (decodificação)

Redes neurais autocodificadoras

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Busca, a partir de uma entrada, reduzi-la (codificação) e, depois, expandi-la novamente (decodificação)
- A base reduzida é a codificação da rede

Redes neurais autocodificadoras

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Busca, a partir de uma entrada, reduzi-la (codificação) e, depois, expandi-la novamente (decodificação)
- A base reduzida é a codificação da rede
- Com essa codificação, podemos utiliza-la para uma classificação

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

Objetivo

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Adaptar uma rede neural autocodificadora para reduzir a quantidade de atributos de uma base de microarranjo

Objetivo

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- Adaptar uma rede neural autocodificadora para reduzir a quantidade de atributos de uma base de microarranjo
- Para isso, serão realizados experimentos em cinco bases de dados.

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não
- Para tal serão utilizados métodos de classificação. São estes:

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não
- Para tal serão utilizados métodos de classificação. São estes:

- 1 KNN

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não
- Para tal serão utilizados métodos de classificação. São estes:
 - 1 KNN
 - 2 SVM

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não
- Para tal serão utilizados métodos de classificação. São estes:
 - 1 KNN
 - 2 SVM
 - 3 Naive Bayes

- Com as novas bases de dados, é necessário fazer uma análise para verificar se essas bases são representativas ou não
- Para tal serão utilizados métodos de classificação. São estes:
 - 1 KNN
 - 2 SVM
 - 3 Naive Bayes
 - 4 Árvore de Decisão

Sumário

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

1 Introdução

2 Técnicas de redução

3 Redes neurais autocodificadoras

4 Objetivo

5 Análise

6 Referências

Redução de Dimensionalidade em Base de Dados de Microarranjos

Rafael Felipe
Tasaka de
Melo
Helyane
Bronoski
Borges

Introdução

Técnicas de redução

Redes neurais autocodificadoras

Objetivo

Análise

Referências

- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**. 2017. O'Reilly Media, Inc. Pag. 206.