

Pre-treinamento de DNN com Deep Belief Networks e Visualização em Tensorboard

Prática

Exemplos

Parte 1: Treinamento de uma DNN

Exemplo 1: MNIST

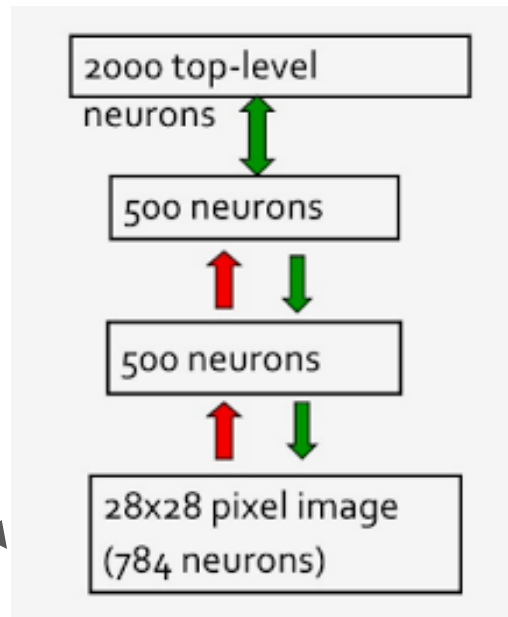
Exemplo: 20 News groups

Parte 2: Visualização e análise de features em Tensorflow

Visualização de Texto

Visualização de Imagens

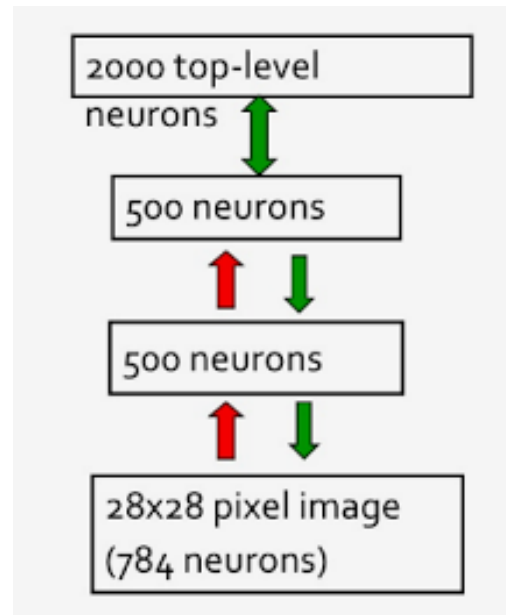
Exemplo 1: MNIST



Exemplo 1: MNIST

Tarefas:

- Configuração apresentada.
- Configuração [1000 1000 1000]
- Treinar a rede sem pre-treinamento
- Treinar a rede com pre-treinamento



Exemplo: 20 News groups

O conjunto de dados 20 Newsgroups é uma coleção de aproximadamente 20.000 documentos de grupo de notícias, particionado (quase) uniformemente em 20 groups diferentes.

Para o melhor de meu conhecimento, ele foi originalmente coletado para o seu Newsweeder: Uma rede que aprender a filtrar notícias.

comp.graphics comp.os.ms-windows.misc comp.sys.ibm.pc.hardware comp.sys.mac.hardware comp.windows.x	rec.autos rec.motorcycles rec.sport.baseball rec.sport.hockey	sci.crypt sci.electronics sci.med sci.space
misc.forsale	talk.politics.misc talk.politics.guns talk.politics.mideast	talk.religion.misc alt.atheism soc.religion.christian



Quem estiver interessado em text mining:

Frameworks:

<http://www.nltk.org/>

<https://spacy.io/>

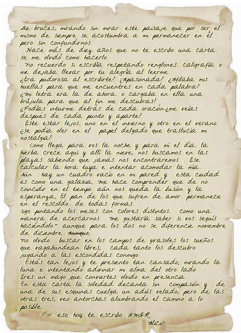
<https://nlp.stanford.edu/>

Tutorial:

<https://www.kaggle.com/c/word2vec-nlp-tutorial#description>

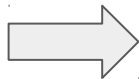
<https://github.com/tensorflow/models/tree/master/syntaxnet>

Como levamos os dados para Deep Learning ?



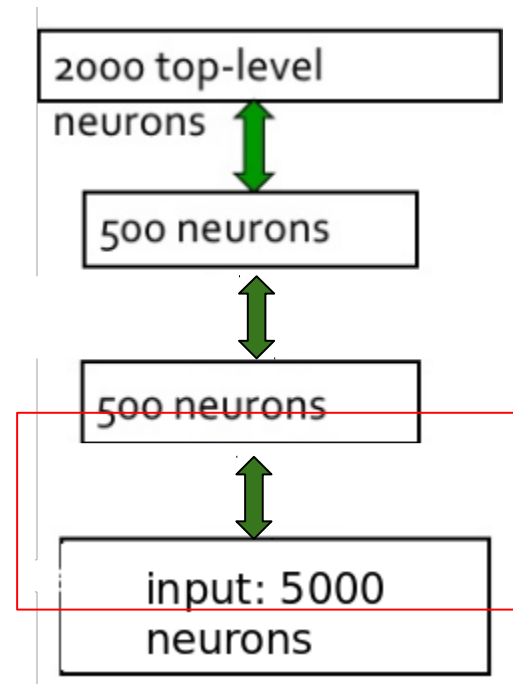
Uma instancia normalmente tem miles de atributos:

0001001011111000000...0000011000000000100000111110

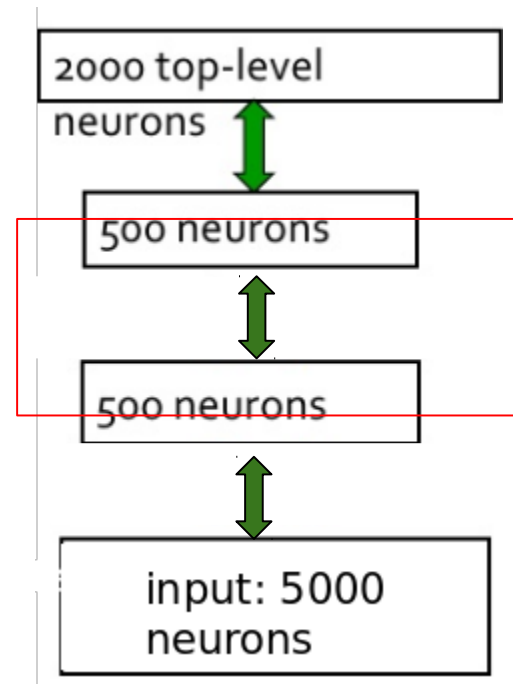


Doc 1: 0001001111111000000...000001100000000010000011111
Doc 2: 0111001011111000000...000001100000000010000011110
Doc 3: 001011011111000000...0010011000010001000001111100
... : 001011011111000000...0000011000000000100000111110
Doc n: 0101101011111000000...000001100000010010000011111

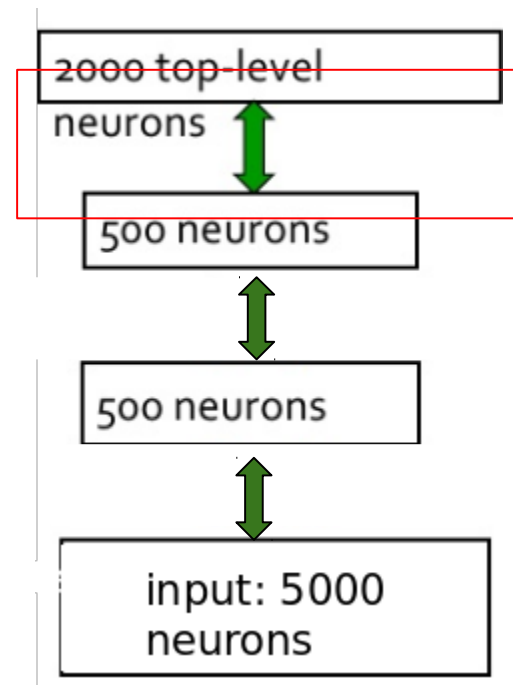
Deep Belief Network



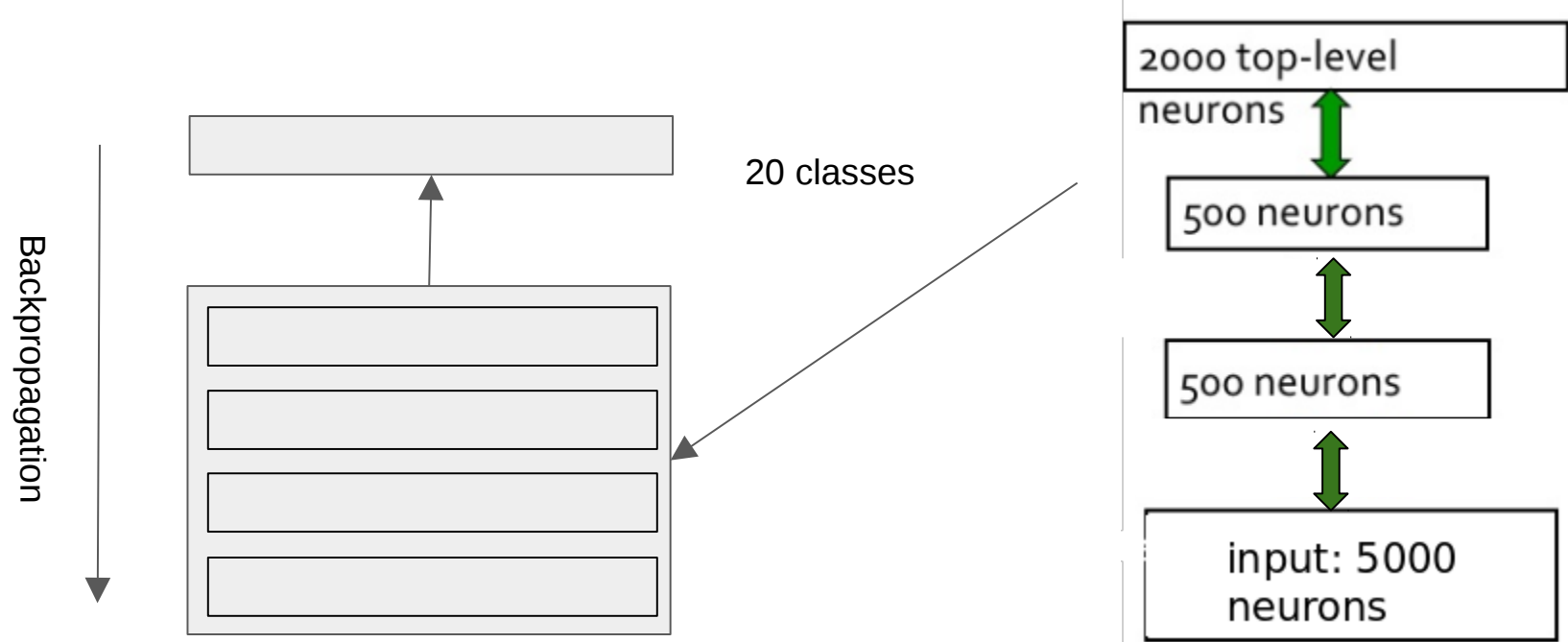
Deep Belief Network



Deep Belief Network



Deep Belief Network

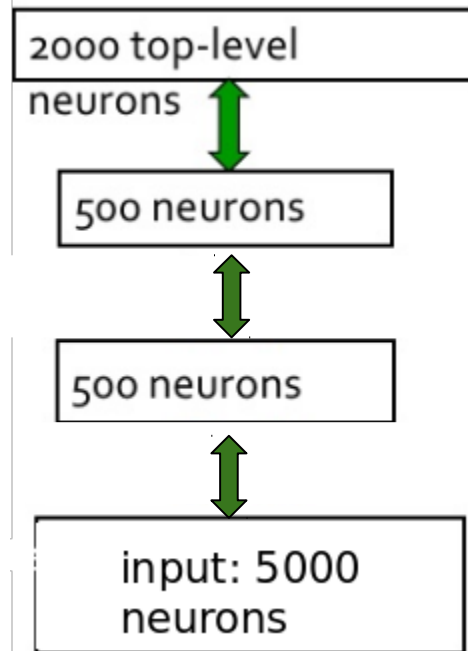


Deep Belief Network

Em geral o processamento de texto é custoso (DL), consome maiores recursos e apresenta maior complexidade.

Tarefas:

- Configuração [1000 1000 1000]
- Treinar a rede sem pre-treinamento
- Treinar a rede com pre-treinamento



Parte 2

Para que utilizar **Embedding Visualization**?

- Além de ser atualmente um dos mais populares visualizadores para altas dimensões, atualmente em DL é direcionada: na customização de Máquinas de Deep Learning em tempo real.
- Modificar a estrutura de dados com que você trabalha aumentando a acurácia.
 - Pode separar os melhores features e levar para treinar um modelo com melhor acurácia.
 - Pode indicar ao modelo reforçar o treinamento em dados que pertencem à mesma classe.
- Indicar similaridade e informar para o algoritmos treinar que é o que tem que melhorar.
- Um artigo que profundiza o analise de esta ferramenta:

Parte 2

Para que utilizar **Embedding Visualization**?

Um artigo que profundiza o analise de esta ferramenta:

- <http://distill.pub/2016/misread-tsne/>

Sites:

- <https://artsexperiments.withgoogle.com/tsnemap/>

Parte 2

Visualização e análise de features em **Tensorflow** (GTC 2017)

Dados que quer
Visualizar

Imagens, Texto, features, ...

Metadados

Rotulos, Imagens, caminhos,

Metadados

Visualização e análise de features em **Tensorflow** (GTC 2017)

Metadados

metadata.tsv

```
a11 a12 a13 a14  
a21 a22 a23 a24  
a31 a32 a33 a34  
a41 a42 a43 a44  
a51 a52 a53 a54  
...
```


Metadados

Visualização e análise de features em **Tensorflow** (GTC 2017)

Dados que quer
Visualizar

