

# Aprendizado de Máquina

## Aprendizado profundo

André C. P. L. F. de Carvalho  
 Pós-doutorando: Isvani Frias-Blanco  
 ICMC-USP

MACHINE LEARNING

# Principais tópicos

- Introdução
- Extração de características
- Redes profundas
- Aprendizado profundo
- Redes autodecodificadoras
- Redes convolucionais

© André de Carvalho - ICMC/USP

2

# Limitações de AM

- Dificuldade de utilizar os dados no formato original
- Necessário extrair características (atributos)
  - Transformar dados de entrada em um vetor de características
    - Que pode ser usado como novo vetor de entrada por algoritmos de AM

© André de Carvalho - ICMC/USP

3

# Extração de características

- Deve gerar representações que:
  - Seleccionam aspectos relevantes para a classificação correta
  - Ignoram aspectos irrelevantes para a classificação
- Alternativa 1: extrair "manualmente" características que considerar relevantes
  - Necessário poder projetar bons extratores e ter bom conhecimento do domínio de aplicação
    - Caro e subjetivo

© André de Carvalho - ICMC/USP

4

# Extração de características

- Alternativa 2: utilizar extrator genérico de características não-lineares
  - Ex.: funções kernel genéricas
    - Gaussiana, polinomial, ...
  - Pode ter baixa generalização para exemplos muito diferentes dos de treinamento
- Alternativa 3: aprender a extrair boas características automaticamente
  - Usando um algoritmo genérico de aprendizado
  - Redes neurais profundas

© André de Carvalho - ICMC/USP

5

# RNs profundas (RNP)

- Redes neurais MLP em geral têm 1 camada intermediária
  - Redes neurais rasas (shallow)
  - Poucas camadas dificulta induzir função que represente bem os dados de treinamento
- Muitas camadas + *backpropagation* levam a soluções pobres
  - Problemas de atribuição de erro e lentidão e
  - Pequena quantidade de exemplos de treinamento

© André de Carvalho - ICMC/USP

6

## Outros problemas do *backprop*

- Pode convergir para mínimos locais
  - Não é tão problemático quando número de atributos preditivos é grande
- Pode usar apenas dados rotulados
  - Maioria dos dados é gerado sem rótulo
  - Custo de rotular manualmente pode ser elevado

© André de Carvalho - ICMC/USP

7

## RNs profundas (RNP)

- Complexidade em teoria de circuitos
  - Sugere que arquiteturas mais profundas podem ser mais eficientes
    - Quando a tarefa é complexa e existem dados suficientes para capturar essa complexidade
      - Big data
    - Necessário utilizar algoritmos de treinamento apropriados

© André de Carvalho - ICMC/USP

8

## RNs profundas (RNP)

- Qualquer algoritmo pode ser implementado em um circuito Booleano profundo o suficiente
  - Uma camada para cada passo de execução do algoritmo
  - Quanto mais profundo o circuito, mais complexo o algoritmo que pode ser implementado

© André de Carvalho - ICMC/USP

9

## RNs profundas (RNP)

- Redes neurais são mais eficientes que circuitos booleanos
  - Precisam de menos camadas que circuitos Booleanos para o mesmo problema
  - Redes mais profundas (RNPs) ⇒
    - Algoritmos com mais passos ⇒
      - Algoritmos mais poderosos

© André de Carvalho - ICMC/USP

10

## RNs profundas (RNP)

- Redes com 2 camadas e poucas unidades escondidas ordenam  $N$  números de  $N$  bits
  - Impossível fazer com circuito Booleano de 2 camadas
    - Precisa de muito mais camadas
  - Neurônios realizam operações de limiar de sinal
- Neurônios naturais são lentos
  - #disparos < 100 vezes por segundo
  - Para resolvermos problema em 0.1s
    - Podemos ter uma sequência de no máximo 10 disparos
    - Rede natural com 10 camadas faz o que fazemos em 0.1s

© André de Carvalho - ICMC/USP

11

## O que podemos fazer em 0.1s

- Reconhecer uma face
- Reconhecer emoção
- Entender uma fala
- Teoricamente, RNPs podem fazer o que seres humanos fazem, em menos tempo
  - Basta encontrar valores corretos para seus pesos



© André de Carvalho - ICMC/USP

12

## Redes profundas

- Arquitetura com várias camadas
  - Mais de uma camada intermediária
  - Camadas podem ter propósitos diferentes
- Aprendizado profundo
  - Algoritmo de treinamento para treinar redes profundas
- Não muda muito em relação às redes MLP e seu treinamento

© André de Carvalho - ICMC/USP

13

## Redes profundas

- Desde as redes MLPs
  - Camadas intermediárias extraem características
  - Dificuldades:
    - Algoritmo de treinamento backpropagation funcionava bem até poucas camadas
      - Problema de atribuição de erro
      - Tecnologia computacional disponível
      - Uso de conjuntos de dados pequenos

© André de Carvalho - ICMC/USP

14

## RNs profundas

- Família de técnicas
  - Características são extraídas hierarquicamente por aprendizado
    - Não supervisionado
      - Pode usar dados não rotulados
    - Semi-supervisionado
  - Características cada vez mais complexas são extraídas

© André de Carvalho - ICMC/USP

15

## Deep learning

- Aprendizado profundo
- Técnicas de aprendizado representacional
  - Converter fala em texto em fala
  - Diagnosticar problemas de acordo com sintomas
  - Identificar objetos em imagens

© André de Carvalho - ICMC/USP

16

## Deep learning

- Aprendem a extrair representações de dados de entrada
  - Que podem ser utilizadas para tarefas de classificação
  - Usam estruturas (arquiteturas) com vários níveis de representação
    - Composição de módulos (camadas) de processamento simples e não linear
      - Cada camada transforma a representação obtida pela camada anterior

© André de Carvalho - ICMC/USP

17

## Extração de características

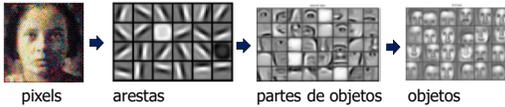
- Representações da entrada se tornam cada vez mais abstratas
- Compor de várias camadas permite extrair características complexas
  - Aprender funções complexas
  - Cada transformação amplifica aspectos importantes para a classificação
    - e elimina variações irrelevantes

© André de Carvalho - ICMC/USP

18

## Extração de características

- Inicialmente características simples
- Nível crescente de abstração
  - Cada camada faz uma transformação não linear das características recebidas



© André de Carvalho - ICMC/USP

19

## Aspectos positivos das RNP

- Biologicamente plausíveis
- Para dados com  $n$  atributos preditivos
  - Número de neurônios necessários em uma rede rasa cresce exponencialmente com  $n$
  - Em uma rede profunda, esse número cresce e no máximo linearmente com  $n$ 
    - Redes mais profundas podem usar menos pesos e parâmetros
    - Ex.: paridade

© André de Carvalho - ICMC/USP

20

## Primeiras redes

- Neognitron, 1980
- Linsker rede neural com conexões laterais, 1986
- MLP com várias camadas, 1989
- Neural history compressor, 1983
- SOFT-GSN, 1984

© André de Carvalho - ICMC/USP

21

## Principais RNs profundas

- Redes credais profundas (RCP)
- Redes autocodificadoras empilhadas (RAE)
- Redes neurais convolucionais (RNC)
- Redes neurais profundas (RNP)

© André de Carvalho - ICMC/USP

22

## Redes neurais profundas

- Redes com mais de uma camada intermediária
  - Em geral, completamente interconectadas
- Geralmente, as camadas intermediárias (ou parte delas) são pré-treinadas
  - Depois, utilizam *backpropagation* nas camadas finais
  - Ou em toda a rede



© André de Carvalho - ICMC/USP

23

## Redes autocodificadoras

- Uma das razões para o ressurgimento do interesse em RNAs
  - Rede MLP em que:
    - Número de neurônios da camada de saída = número de entradas
    - Uma camada intermediária com muito menos neurônios que a camada de saída
    - Treinamento ensina a rede a responder com um vetor de saída igual ao vetor de entrada
    - Criadas na década de 1980

© André de Carvalho - ICMC/USP

24

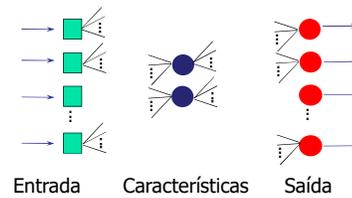
## Redes autocodificadoras

- Rede aprende a codificar a entrada usando poucos bits
  - Camada de saída funciona como decodificador para o tamanho original
  - Quanto menos neurônios na camada intermediária, menor o tamanho do código
    - Semelhante a programas de compressão de arquivos, mas a rede aprende como comprimir

© André de Carvalho - ICMC/USP

25

## Rede autocodificadora



© André de Carvalho - ICMC/USP

26

## Redes autocodificadoras

- Poucos neurônios na camada intermediária dificultavam aprendizado
  - Tinham que codificar uma grande quantidade de informação em poucos bits
    - Saíram de moda
  - Cerca de uma década depois uma nova proposta:
    - Usar uma camada intermediária maior que as camadas de entrada e de saída

© André de Carvalho - ICMC/USP

27

## Redes autocodificadoras esparsas

- Cada ajuste dos pesos ignora a maioria dos neurônios da camada intermediária
  - Camada escondida não copia entrada
  - Aprendizado de torna bem mais fácil
- Mais adiante, foi proposto o uso de várias camadas codificadoras
  - Redes autocodificadoras empilhadas

© André de Carvalho - ICMC/USP

28

## Rede autocodificadora empilhada

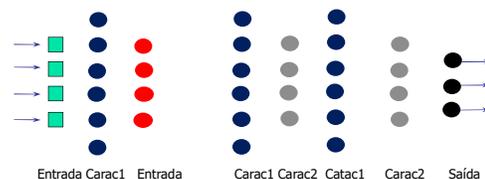
- Empilha autodecodificadores esparsos
  - Camada escondida do primeiro autodecodificador, se torna entrada e saída do segundo
  - Camada escondida do segundo, se torna entrada e saída do terceiro ...
  - Cada camada intermediária aprende uma representação mais sofisticada da entrada
    - Camada intermediária anterior

© André de Carvalho - ICMC/USP

29

## Rede autocodificadora empilhada

- Treina autodecodificadores sequencialmente
  - Abordagem gulosa de treinamento
  - Altera camada de entrada e de saída a cada vez



© André de Carvalho - ICMC/USP

30

## Rede autocodificadora empilhada

- Última camada da (camada de saída)
  - Classificador softmax
    - Generalização da função logística
    - Transforma um vetor com  $k$  valores reais em  $k$  valores  $\in (0,0,1,0]$  cuja soma = 1
  - Treinamento
    - Indução de classificador
    - Treina última camada (ou toda a rede) com *backpropagation*

© André de Carvalho - ICMC/USP

31

## RNs convolucionais

- ConvNets ou CNN
- Propostas para visão computacional
  - Procuram padrões visuais utilizando os pixels das imagens
  - Robustas a distorções e transformações geométricas
- Semelhante a rede *neocognitron*, proposta na década de 80

© André de Carvalho - ICMC/USP

32

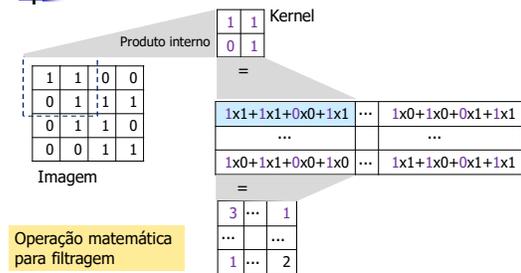
## RNs convolucionais

- Utilizam camadas para convolução de uma imagem
  - Camadas de convolução aplicam transformações a pixels de uma imagem de entrada
- Compostas por dois estágios
  - Sequência de pares de camadas
    - Camada convolucional
      - Extrai mapas de características de uma imagem usando filtros
    - Camada de amostragem (pooling)
      - Mantém apenas as informações mais relevantes dos mapas
  - Rede MLP convencional

© André de Carvalho - ICMC/USP

33

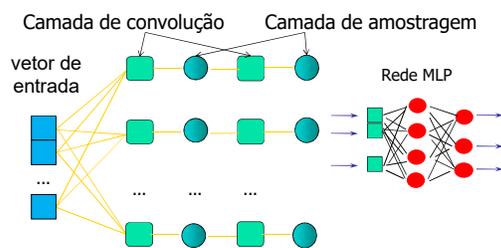
## Convolução



© André de Carvalho - ICMC/USP

34

## RNs convolucionais



© André de Carvalho - ICMC/USP

35

## Redes credais profundas

- Modelos probabilísticos
- Treinamento guloso camada a camada
  - Primeiras camadas recebem conexão top down da camada seguinte
    - Cada camada é uma Máquina de Boltzman restrita
  - Duas últimas camadas têm conexões bidirecionadas
    - Com pesos simétricos entre elas

© André de Carvalho - ICMC/USP

36

## Conclusão

- Introdução
- Extração de características
- Redes profundas
- Aprendizado profundo
- Redes autodecodificadoras
- Redes convolucionais

© André de Carvalho - ICMC/USP 37

## Perguntas



© André de Carvalho - ICMC/USP 38