

Conceitos e aplicações da aprendizagem de máquina

Eduardo Krempser
eduardo.krempser@fiocruz.br

Matheus Müller
matheusp@posgrad.Incc.br

Atributos

Definição de Atributos

objetos, registros, amostras, ...

| Localização | Quartos | Garagem | ... | Aluguel (R\$/m ²) |
|-----------------|---------|---------|-----|-------------------------------|
| Copacabana | 2 | Não | ... | 41,96 |
| Copacabana | 1 | Sim | ... | 49,15 |
| Centro | 2 | Não | ... | 28,05 |
| Barra da Tijuca | 3 | Sim | ... | 30,86 |
| ... | ... | ... | ... | ... |

alvo, classe, ...

atributos, variáveis, descritores, ...

Definição de Atributos

- Tipos de atributos:
 - Nominal
 - ex.: identificadores, cor dos olhos, CEP
 - Ordinal
 - ex.: Ordenamentos (sabor de um prato em uma escala de 0 a 10), notas
 - Intervalo
 - ex.: Intervalo entre datas, temperatura em Celsius
 - Taxas
 - Valor monetário, massa, comprimento

Definição de Atributos

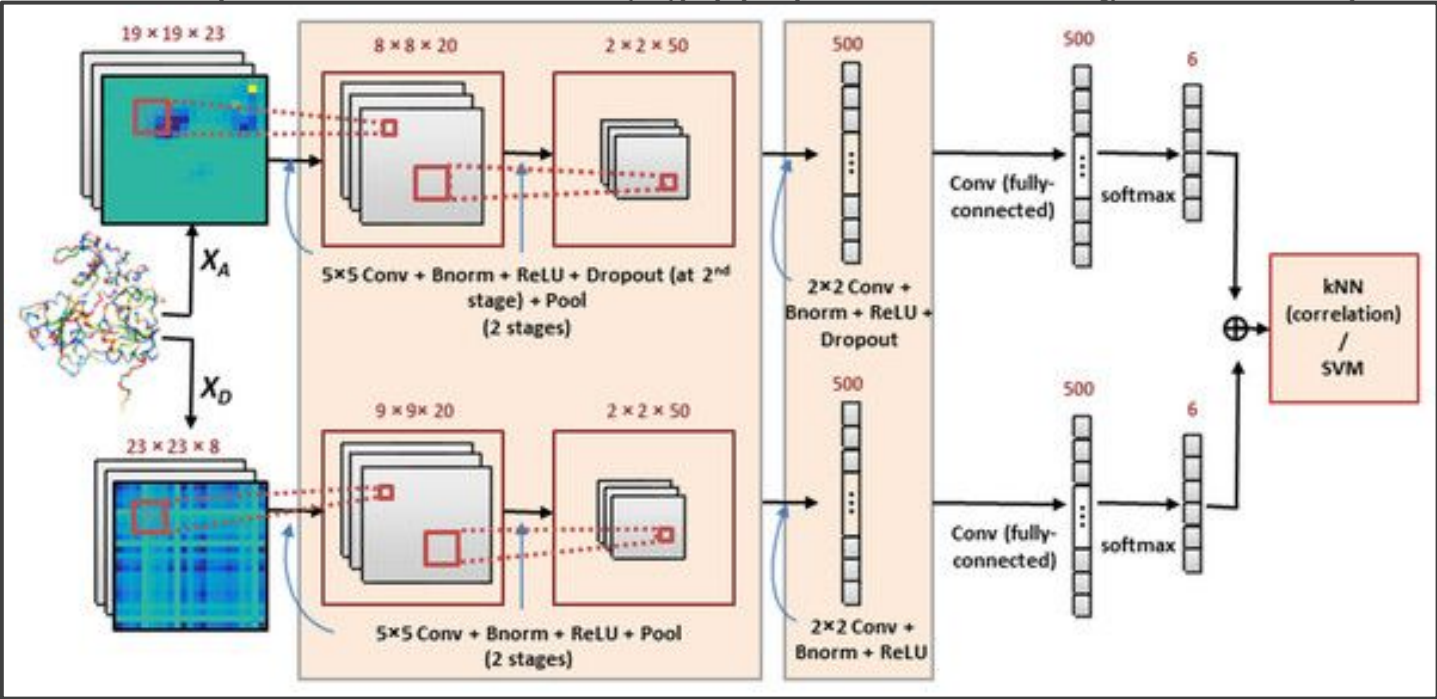
- Discretos
 - Possui um finito ou contável infinito conjunto de valores
 - Frequentemente representado por variáveis
 - Atributos binários são casos especiais de atributos discretos
- Contínuos
 - Possui seus valores reais
 - Na prática, valores reais só podem ser medidos e representados usando um conjunto finito de dígitos
 - Tipicamente representado por variáveis do tipo “float”

Definição de Atributos

- Estamos preocupados com a representatividade e as operações que podem ser aplicadas
 - $= \neq$
 - $< >$
 - $* /$
 - $+ -$

Definição de Atributos

- Estamos ser aplica
 - \neq
 - $< >$
 - $* /$
 - $+ -$

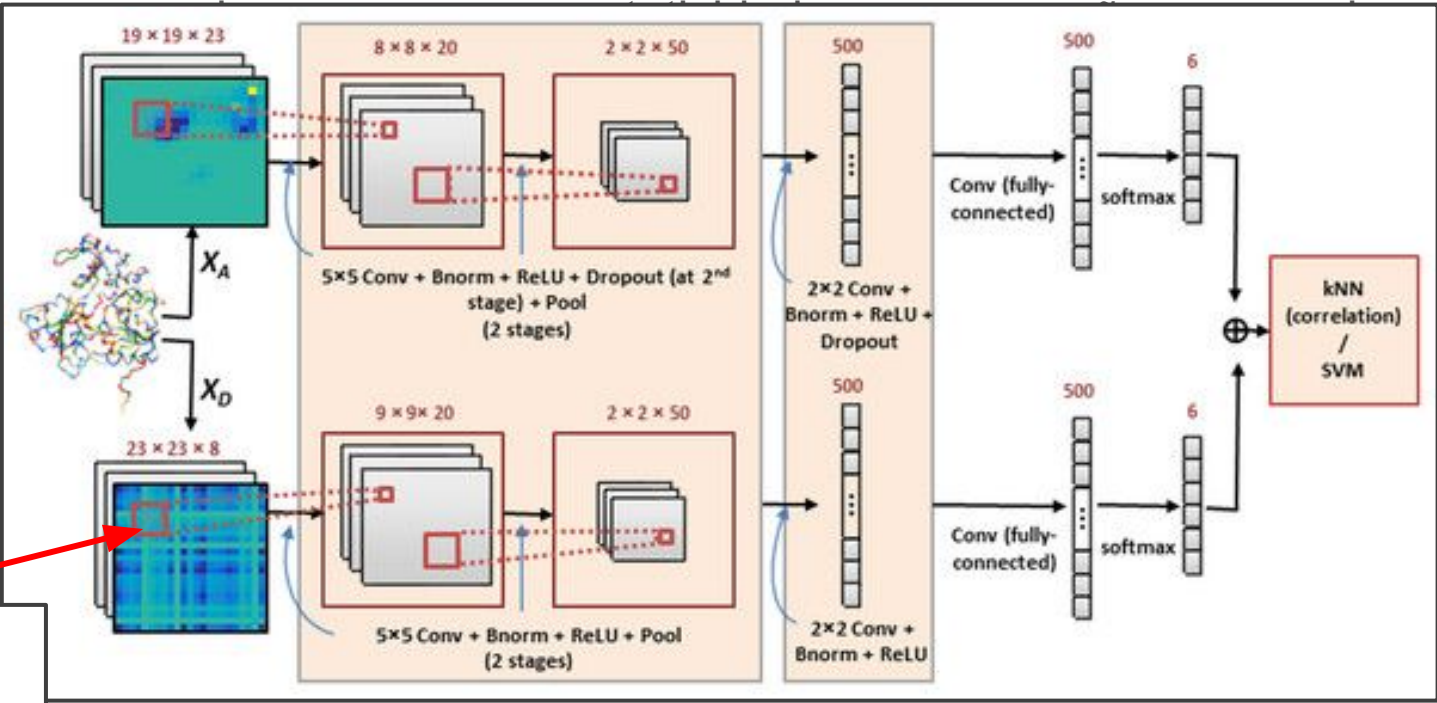


Prediction of protein function using a deep convolutional neural network ensemble. Zacharaki, E. I, 2017

Definição de Atributos

- Estamos
ser aplica

- $\circ = \neq$
- $\circ < >$
- $\circ * /$
- $\circ + -$



[0, 255] ?

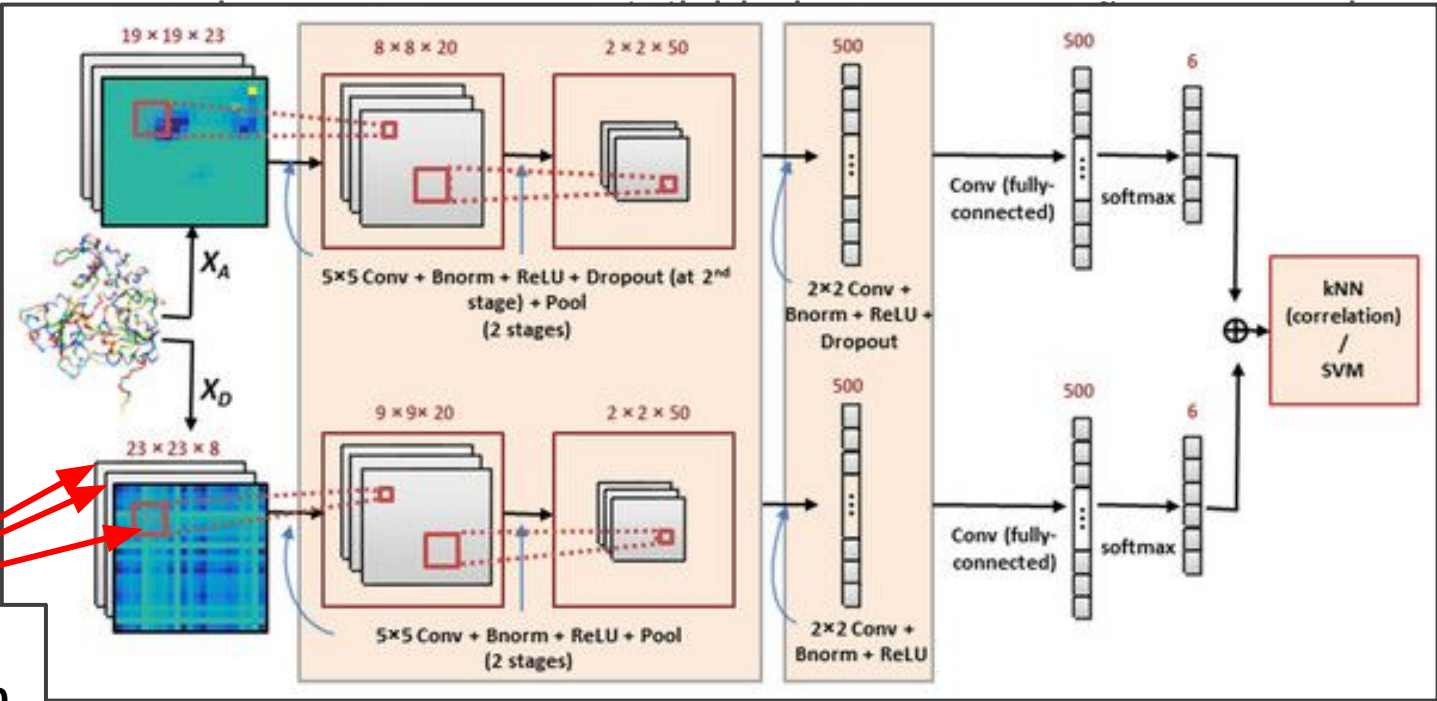
Prediction of protein function using a deep convolutional neural network ensemble. Zacharaki, E. I, 2017

Definição de Atributos

- Estamos ser aplica

- $\circ = \neq$
- $\circ < >$
- $\circ * /$
- $\circ + -$

$[0, 255], [0, 255], [0, 255]$?



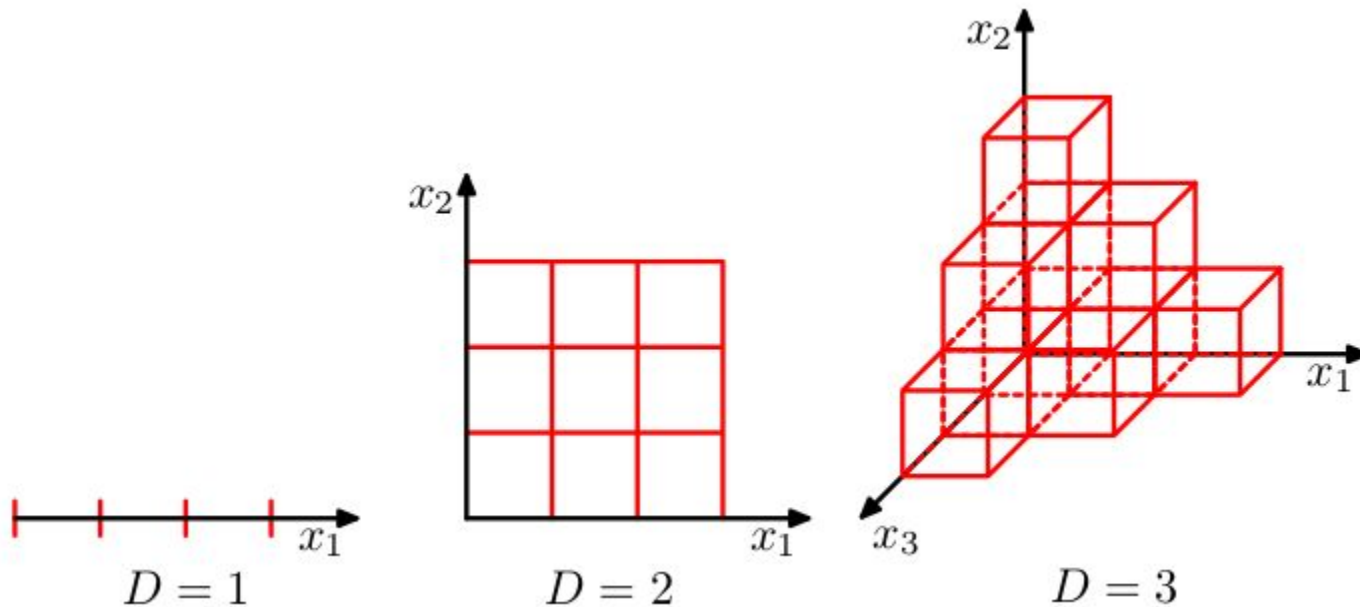
Prediction of protein function using a deep convolutional neural network ensemble. Zacharaki, E. I, 2017

Características dos Dados

- Dimensionalidade (número de atributos)
 - Alta dimensionalidade geram uma série de desafios
- Esparsidade
 - Dados de presença, por exemplo
- Resolução
 - Padrões dependem da escala
- Tamanho
 - Tipos de análises podem depender do tamanho do conjunto de dados

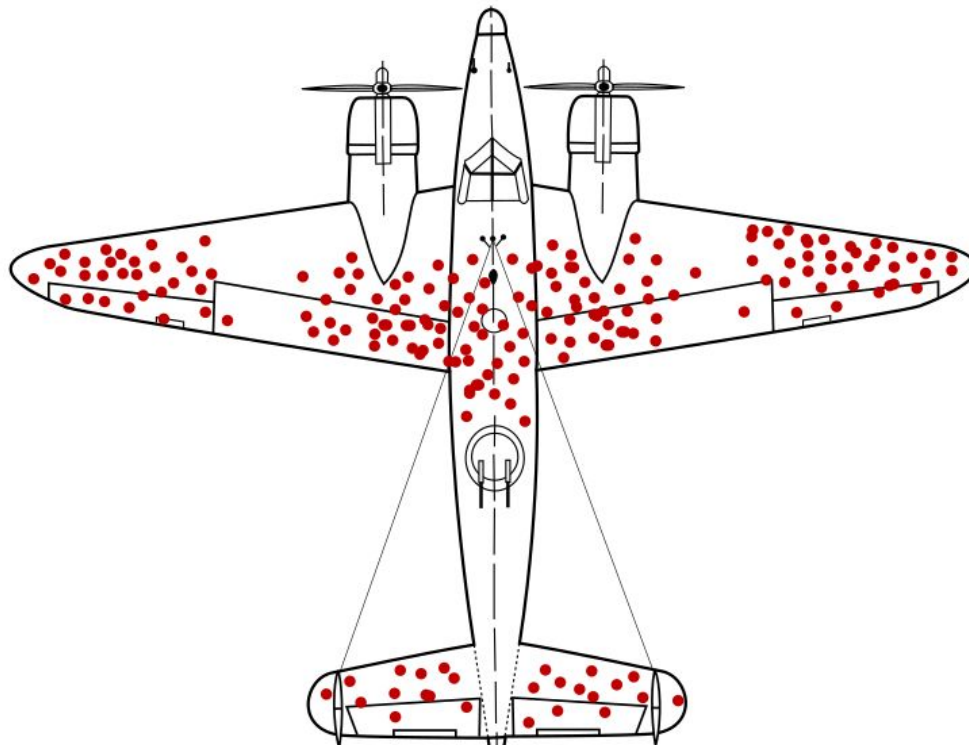
Fontes de dados

“Maldição” da Dimensionalidade



Pattern Recognition and Machine Learning. Bishop, C. M.; 2006

Viés do Sobrevivente



<https://pt.wikipedia.org/wiki/Ficheiro:Survivorship-bias.svg>

Fonte de Dados - Prática

- Aula 02 - Polinômio
- Aula 02 - Íris
- Aula 02 - QSAR - Biodegradável

Conceitos e aplicações da aprendizagem de máquina

Eduardo Krempser
eduardo.krempser@fiocruz.br

Matheus Müller
matheus.mullerps@gmail.com