

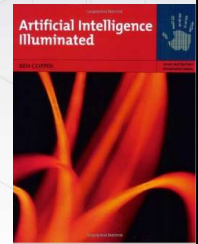
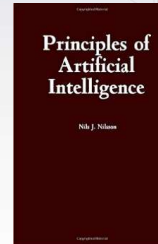
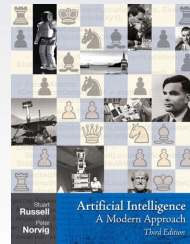
Inteligência Artificial

Sistemas de Informação

Cleiton Garcia <cleitons@catolicasc.org.br>



Referências



13/03/2018

Prof. Msc. Cleiton Garcia

2



Revisão...

- Aprendizagem supervisionada
 - Classificação e Regressão
 - Treinamento e Testes
 - Medições da acurácia e erro
- Sobre ou Sub-ajuste
- Árvore de decisão
 - Implementação utilizada p/ Classificação e Regressão?

13/03/2018

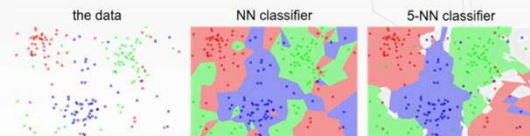
Prof. Msc. Cleiton Garcia

3



Aprendizagem baseada em instâncias

- Categoria de aprendizagem muito utilizada
- São métodos não paramétricos/normal: podem ser usados com distribuições arbitrárias e sem a suposição de que a forma das densidades são conhecidas.
- Armazenam os exemplos de treinamento.
- A generalização é feita somente quando uma nova instância tiver que ser classificada.



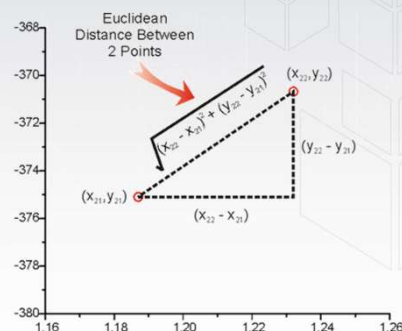
13/03/2018

Prof. Msc. Cleiton Garcia

4



Medidas de distância - Euclidiana



13/03/2018

Prof. Msc. Cleiton Garcia

5



Medidas de distância - Euclidiana

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

- Exemplo em 3 dimensões
- ```
private Double distance(RGB o1, RGB o2) {
 return Math.sqrt(
 Math.pow((o2.R - o1.R), 2)
 + Math.pow((o2.G - o1.G), 2)
 + Math.pow((o2.B - o1.B), 2));
}
```

13/03/2018

Prof. Msc. Cleiton Garcia

6



### Outras funções usadas para distância

|                                |                                                                                                                                                                                                                                     |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Euclidean                      | $d(x,y) = \sqrt{\sum (x_i - y_i)^2}$                                                                                                                                                                                                |
| Squared Euclidean              | $d(x,y) = \sum (x_i - y_i)^2$                                                                                                                                                                                                       |
| Manhattan                      | $d(x,y) = \sum  x_i - y_i $                                                                                                                                                                                                         |
| Canberra                       | $d(x,y) = \sum \frac{ x_i - y_i }{x_i + y_i}$                                                                                                                                                                                       |
| Chebyshev                      | $d(x,y) = \max  x_i - y_i $                                                                                                                                                                                                         |
| Bray Curtis                    | $d(x,y) = \frac{\sum  x_i - y_i }{\sum x_i + \sum y_i}$                                                                                                                                                                             |
| Cosine                         | $d(x,y) = \frac{\sum (x_i y_i)}{\sqrt{\sum x_i^2 \sum y_i^2}}$                                                                                                                                                                      |
| Pearson Correlation            | $d(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$                                                                                                                         |
| Uncentered Pearson Correlation | $d(x,y) = \frac{\sum x_i y_i}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$                                                                                                                                                |
| Euclidean Nullweighted         | Same as Euclidean, but only the indexes where both x and y have a value (not NULL) are used, and the result is weighted by the number of values calculated. Nulls must be replaced by the missing value calculator (in dataloader). |

13/03/2018

Prof. Msc. Cleiton Garcia

7



### K-Nearest Neighbor (K-NN)

- Classificar  $x_t$  atribuindo a ele o rótulo representado mais frequentemente dentre as k amostras mais próximas e utilizando um esquema de votação.

13/03/2018

Prof. Msc. Cleiton Garcia

8



### K-Nearest Neighbor (K-NN)

- K - vizinhos mais próximos é um algoritmo de classificação (ou regressão).
- É supervisionado porque você está tentando classificar um ponto com base na classificação conhecida de outros pontos.
- O classificador baseia-se na distância dos K - vizinhos mais próximos
- O novo objeto é classificado baseado no treinamento dos atributos das amostras
- A classificação é por meio de voto majoritário entre os K objetos de uma classes (combina a classificação)

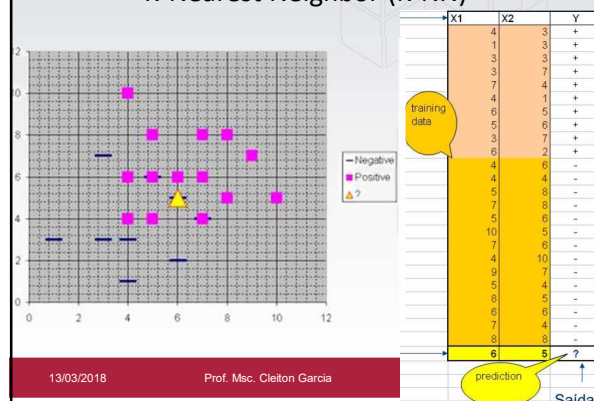
13/03/2018

Prof. Msc. Cleiton Garcia

9



### K-Nearest Neighbor (K-NN)



13/03/2018

Prof. Msc. Cleiton Garcia

### K-NN passo a passo

- Definir o valor de K como parâmetro do algoritmo.
- Calcular a distância (Euclidiana) entre a amostra desconhecida e as de treinamento.
- Ordenar as distâncias e buscar pelos k-vizinhos mais próximos ou k-menores distâncias.
- Utilizar o voto majoritário entre as k-vizinhos mais próximos para prever o valor de amostra desconhecida.

13/03/2018

Prof. Msc. Cleiton Garcia

11



### K-NN passo a passo

- Para K = 8 (nós utilizamos os 8 vizinhos mais próximos)
- Resultado do voto majoritário
- Y = Vm (X5, X7, X8, X15, X17, X20, X22, X23)
- \* Vm = Voto majoritário

| Computation |                  |      |
|-------------|------------------|------|
| Distance    | Nearest Neighbor | sign |
| 8           |                  |      |
| 29          |                  |      |
| 13          |                  |      |
| 13          |                  |      |
| 2           | +                |      |
| 20          |                  |      |
| 0           | +                |      |
| 2           | +                |      |
| 13          |                  |      |
| 9           |                  |      |
| 5           |                  |      |
| 5           |                  |      |
| 10          |                  |      |
| 10          |                  |      |
| 2           | -                |      |
| 16          |                  |      |
| 2           | -                |      |
| 29          |                  |      |
| 13          |                  |      |
| 2           | -                |      |
| 4           |                  |      |
| 1           | -                |      |
| 2           | -                |      |
| 13          |                  |      |

13/03/2018

Prof. Msc. Cleiton Garcia

## K-NN

- Vantagens
  - Robusto contra ruídos no conjunto de dados de treinamento
  - Efetivo quando o conjunto de dados de treinamento é elevado
  - A informação presente nos exemplos de treinamento nunca é perdida.
- Desvantagens
  - Necessário definir o valor de K
  - Custo computacional elevado pois é necessário calcular as distâncias entre todos os dados de treinamento, para cada nova predição. Quanto mais exemplos de treinamento maior o custo.

13/03/2018

Prof. Msc. Cleiton Garcia

13



## O que é similaridade?

- A qualidade, caráter ou condição das coisas similares (*Dicionário Houaiss*) .
- Similaridade é difícil de definir, mas...  
*Reconhecemos quando a vemos!*



13/03/2018

Prof. Msc. Cleiton Garcia

14



## Resumo

- Métodos de aprendizagem baseados em instâncias não necessitam formar uma hipótese explícita da função alvo sobre o espaço das instâncias.
- Eles formam uma aproximação local da função alvo para cada nova instância a "classificar".
- O k-NN é um algoritmo baseado em instâncias para aproximar funções alvo de valor real ou de valor discreto, assumindo que as instâncias correspondem a pontos em um espaço d-dimensional
- O valor da função alvo para um novo ponto é estimada a partir dos valores conhecidos dos k exemplos de treinamento mais próximos.

13/03/2018

Prof. Msc. Cleiton Garcia

15



## Exercício – Desenvolver algoritmo K-nn

1. Dado o conjunto de treinamento abaixo, referente aos valores RGB de uma imagem:

| R   | G   | B   | classes |
|-----|-----|-----|---------|
| 1   | 10  | 200 | 1       |
| 2   | 20  | 230 | 1       |
| 6   | 25  | 150 | 1       |
| 7   | 45  | 100 | 1       |
| 10  | 50  | 125 | 1       |
| 3   | 24  | 111 | 1       |
| 100 | 4   | 10  | 2       |
| 250 | 7   | 50  | 2       |
| 243 | 5   | 68  | 2       |
| 210 | 2   | 90  | 2       |
| 200 | 1   | 95  | 2       |
| 215 | 0   | 68  | 2       |
| 56  | 200 | 1   | 3       |
| 79  | 234 | 3   | 3       |
| 80  | 210 | 8   | 3       |
| 95  | 200 | 10  | 3       |
| 80  | 210 | 4   | 3       |
| 49  | 207 | 1   | 3       |

Efetuar a classificação da base de teste abaixo para valores de k = 3, 5 e 7, para os modelo Um-Contratodos.

| R   | G   | B   | classes |
|-----|-----|-----|---------|
| 1   | 2   | 100 | ?       |
| 10  | 20  | 30  | ?       |
| 8   | 5   | 20  | ?       |
| 237 | 45  | 100 | ?       |
| 1   | 50  | 101 | ?       |
| 67  | 121 | 12  | ?       |

13/03/2018

Prof. Msc. Cleiton Garcia

16



## Exercício – Utilizar IBK do WEKA

1. Dado o conjunto de treinamento abaixo, referente aos valores RGB de uma imagem:

| R   | G   | B   | classes |
|-----|-----|-----|---------|
| 1   | 10  | 200 | 1       |
| 2   | 20  | 230 | 1       |
| 6   | 25  | 150 | 1       |
| 7   | 45  | 100 | 1       |
| 10  | 50  | 125 | 1       |
| 3   | 24  | 111 | 1       |
| 100 | 4   | 10  | 2       |
| 250 | 7   | 50  | 2       |
| 243 | 5   | 68  | 2       |
| 210 | 2   | 90  | 2       |
| 200 | 1   | 95  | 2       |
| 215 | 0   | 68  | 2       |
| 56  | 200 | 1   | 3       |
| 79  | 234 | 3   | 3       |
| 80  | 210 | 8   | 3       |
| 95  | 200 | 10  | 3       |
| 80  | 210 | 4   | 3       |
| 49  | 207 | 1   | 3       |

Efetuar a classificação da base de teste abaixo para valores de k = 3, 5 e 7, para os modelo Um-Contratodos.

| R   | G   | B   | classes |
|-----|-----|-----|---------|
| 1   | 2   | 100 | ?       |
| 10  | 20  | 30  | ?       |
| 8   | 5   | 20  | ?       |
| 237 | 45  | 100 | ?       |
| 1   | 50  | 101 | ?       |
| 67  | 121 | 12  | ?       |

13/03/2018

Prof. Msc. Cleiton Garcia

17



## Seminário

- Elaborar resumo do seu seminário deve ter:
  - ~2 frases de contextualização
  - ~1 frase explicando o objetivo do trabalho
  - ~1 frase sobre o método (como)
  - ~1 frase sobre os resultados esperados
  - ~2 frases com a conclusão e a importância da contribuição ou relevância que motivou vocês a elaborarem este trabalho.
- Submeter no AVA

13/03/2018

Prof. Msc. Cleiton Garcia

18



### Próxima aula

- Trabalhar na elaboração do seminário
- Lembrem-se da dica da continuidade para o trabalho final
- Seminário:
  - artigo 3 páginas no template padrão (SBC)
  - códigos de exemplo
  - Participação (questões na avaliação)
  - Mínimo de 1 referência de artigo científico

13/03/2018

Prof. Msc. Cleiton Garcia

19

