

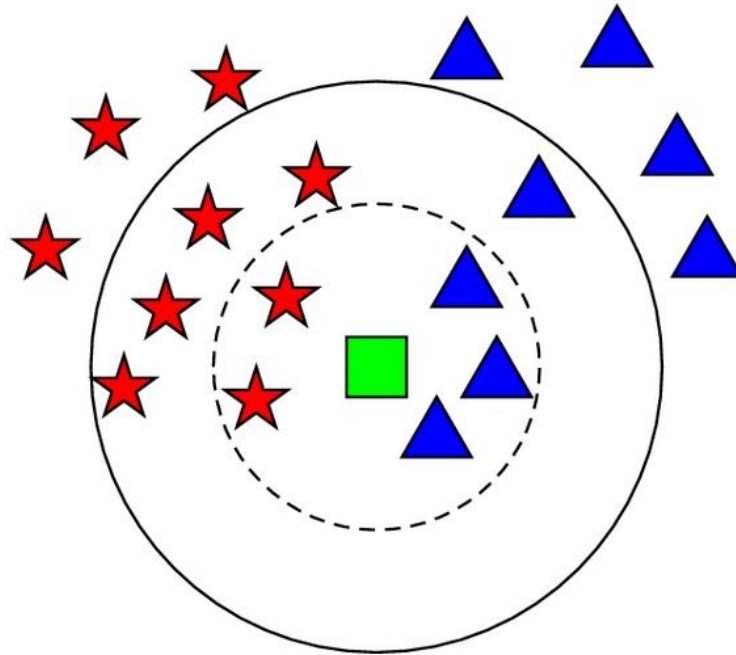
Aprendizado de Máquina

Knn e Kmeans

Felipe Zarattini / 1220557
Pedro de Andrade / 1221273

Knn

Conceito



Knn

Características

- ▶ Fácil implementação
- ▶ Muito efetivo se há muito dados de treinamento
- ▶ É necessário descobrir um valor bom para K

Representação dos Dados - Knn

```
//criando a classe flor
class Flor
{
    //variaveis
private:
    string classe;
    double arg1, arg2, arg3, arg4; //parametros das flores + tipo

public:
    //Construtor
    Flor(double arg1, double arg2, double arg3, double arg4, string classe)
    {
        this->arg1 = arg1;
        this->arg2 = arg2;
        this->arg3 = arg3;
        this->arg4 = arg4;
        this->classe = classe;
    }
}
```

```
//Getters
double getArg1()
{
    return arg1;
}

double getArg2()
{
    return arg2;
}

double getArg3()
{
    return arg3;
}

double getArg4()
{
    return arg4;
}

string getClasse()
{
    return classe;
}

};
```

Resultados Obtidos - Knn

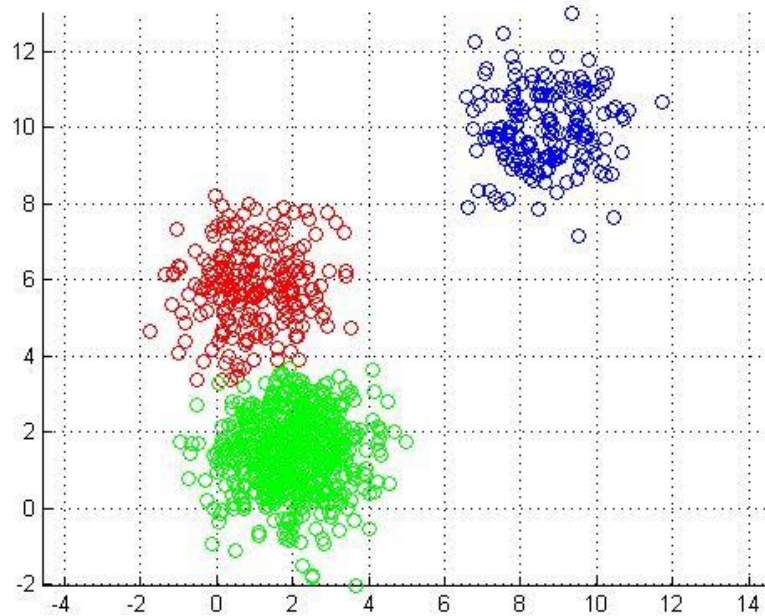
```
Tempo passado para calcular o processo de classificacao: 0.129  
Valores utilizados: K = 3  
Quantidade de Exemplos de Teste = 60  
86 acertos de um total de 90 testes.  
A porcentagem de classificacoes corretas foi de: 95%
```

```
Tempo passado para calcular o processo de classificacao: 0.151  
Valores utilizados: K = 11  
Quantidade de Exemplos de Teste = 50  
98 acertos de um total de 100 testes.  
A porcentagem de classificacoes corretas foi de: 98%
```

```
Tempo passado para calcular o processo de classificacao: 0.029  
Valores utilizados: K = 51  
Quantidade de Exemplos de Teste = 130  
20 acertos de um total de 20 testes.  
A porcentagem de classificacoes corretas foi de: 100%
```

Kmeans

Conceito



Kmeans

Características

- ▶ Método não supervisionado mais simples de implementar
- ▶ Todos os pontos são atribuídos automaticamente a um grupo
- ▶ Valor de K deve ser pré-definida
- ▶ Todos os pontos são forçados a pertencer a um grupo

Representação dos Dados - Kmeans

```
class Ponto
{
private:
    //atributos da classe
    int qtdAtributos, idPonto, idCluster;
    vector<double> atributos;
    string nome;

public:
    //construtor
    Ponto(int idPonto, vector<double>& atributos, string nome = "")
    {
        this->idPonto = idPonto;
        qtdAtributos = atributos.size();

        for(int i = 0; i < qtdAtributos; i++)
            this->atributos.push_back(atributos[i]);

        this->nome = nome;
        idCluster = -1;
    }
};
```

```
//retorna o ID do ponto em questão
int getID()
{
    return idPonto;
}

//retorna a classe do ponto em questão
string getNome()
{
    return nome;
}

//retorna um atributo do ponto em questão a partir de seu indice
double getAtributo(int index)
{
    return atributos[index];
}

//retorna o ID do cluster ao qual o ponto pertence
int getIdCluster()
{
    return idCluster;
}

//retorna a quantidade de atributos do ponto
int getQtdAtributos()
{
    return qtdAtributos;
}

//determina o ID do cluster em qual o ponto está sendo inserido
void setIdCluster(int idCluster)
{
    this->idCluster = idCluster;
}

};
```


Representação dos Dados - Kmeans

```
class Cluster
{
private:
    //atributos da classe
    int idCluster;
    vector<double> centroides;
    vector<Ponto> pontos;

public:
    //construtor
    Cluster(int idCluster, Ponto ponto)
    {
        this->idCluster = idCluster;

        int qtdAtributos = ponto.getQtdAtributos();

        for(int i = 0; i < qtdAtributos; i++)
            centroides.push_back(ponto.getAtributo(i));

        pontos.push_back(ponto);
    }
}
```

```
//retorna o ID do cluster em questao
int getID()
{
    return idCluster;
}

//retorna um ponto inserido no cluster em questão a partir de seu indice
Ponto getPonto(int index)
{
    return pontos[index];
}

//retorna a quantidade de pontos inseridos no cluster em questao
int getQtdPontos()
{
    return pontos.size();
}

//retorna o centroide do cluster em questao a partir de seu indice
double getCentroide(int index)
{
    return centroides[index];
}

//insere um ponto no cluster em questao passado o ponto
void inserePonto(Ponto ponto)
{
    pontos.push_back(ponto);
}

//determina o atributo que servirá de centroide do cluster em questao
void setCentroide(int index, double atributo)
{
    centroides[index] = atributo;
}

//remove um ponto do cluster em questao passado o id do ponto
bool removePonto(int idPonto)
{
    int qtdPontos = pontos.size();

    for(int i = 0; i < qtdPontos; i++)
    {
        if(pontos[i].getID() == idPonto)
        {
            pontos.erase(pontos.begin() + i);
            return true;
        }
    }
}
```

Representação dos Dados - Kmeans

```
class KMeans
{
public:
    //construtor
    KMeans(int K, int qtdPontos, int qtdAtributos, int maxIter)
    {
        this->K = K;
        this->qtdPontos = qtdPontos;
        this->qtdAtributos = qtdAtributos;
        this->maxIter = maxIter;
    }
}
```

```
//processa os pontos da amostra e associa cada ponto ao centroide mais proximo
void processa(vector<Ponto> & pontos)
```

```
//le o arquivo de saida para verificar a porcentagem de pontos que foram
//inseridos incorretamente em cada cluster
void verifica(int K, int qtdPontos)
{
```

Resultados Obtidos - Kmeans

```
>kmeans.exe 3 < dataset.txt
```

Cluster 1

Ponto 108: 7.3 2.9 6.3 1.8 - Iris-virginica
Ponto 101: 6.3 3.3 6.2 5 - Iris-virginica
Ponto 103: 7.1 3 5.9 2.1 - Iris-virginica
Ponto 105: 6.5 3 5.8 2.2 - Iris-virginica
Ponto 106: 7.6 3 6.6 2.1 - Iris-virginica
Ponto 109: 6.7 2.5 5.8 1.8 - Iris-virginica
Ponto 110: 7.2 3.6 6.1 2.5 - Iris-virginica
Ponto 118: 7.7 3.8 6.7 2.2 - Iris-virginica
Ponto 119: 7.7 2.6 6.9 2.3 - Iris-virginica
Ponto 121: 6.9 3.2 5.7 2.3 - Iris-virginica
Ponto 123: 7.7 2.8 6.7 2 - Iris-virginica
Ponto 125: 6.7 3.3 5.7 2.1 - Iris-virginica
Ponto 126: 7.2 3.2 6 1.8 - Iris-virginica
Ponto 130: 7.2 3 5.8 1.6 - Iris-virginica
Ponto 131: 7.4 2.8 6.1 1.9 - Iris-virginica
Ponto 132: 7.9 3.8 6.4 2 - Iris-virginica
Ponto 136: 7.7 3 6.1 2.3 - Iris-virginica
Ponto 141: 6.7 3.1 5.6 2.4 - Iris-virginica
Ponto 144: 6.8 3.2 5.9 2.3 - Iris-virginica
Ponto 145: 6.7 3.3 5.7 2.5 - Iris-virginica
Ponto 113: 6.8 3 5.5 2.1 - Iris-virginica
Ponto 140: 6.9 3.1 5.4 2.1 - Iris-virginica
Ponto 117: 6.5 3 5.5 1.8 - Iris-virginica
Ponto 129: 6.4 2.8 5.6 2.1 - Iris-virginica
Ponto 133: 6.4 2.8 5.6 2.2 - Iris-virginica
Ponto 137: 6.3 3.4 5.6 2.4 - Iris-virginica
Ponto 142: 6.9 3.1 5.1 2.3 - Iris-virginica
Ponto 146: 6.7 3 5.2 2.3 - Iris-virginica
Ponto 104: 6.3 2.9 5.6 1.8 - Iris-virginica
Ponto 111: 6.5 3.2 5.1 2 - Iris-virginica
Ponto 116: 6.4 3.2 5.3 2.3 - Iris-virginica
Ponto 138: 6.4 3.1 5.5 1.8 - Iris-virginica
Ponto 148: 6.5 3 5.2 2 - Iris-virginica
Ponto 149: 6.2 3.4 5.4 2.3 - Iris-virginica
Ponto 53: 6.9 3.1 4.9 1.5 - Iris-versicolor
Ponto 78: 6.7 3 5 1.7 - Iris-versicolor
Ponto 112: 6.4 2.7 5.3 1.9 - Iris-virginica
Ponto 135: 6.1 2.6 5.6 1.4 - Iris-virginica
Atributos do Cluster: 6.85 3.07368 5.74211 2.07105

Cluster 2

Ponto 1: 5.1 3.5 1.4 0.2 - Iris-setosa
Ponto 2: 4.9 3 1.4 0.2 - Iris-setosa
Ponto 3: 4.7 3.2 1.3 0.2 - Iris-setosa
Ponto 4: 4.6 3.1 1.5 0.2 - Iris-setosa
Ponto 5: 5.3 6.1 4 0.2 - Iris-setosa
Ponto 6: 5.4 3.9 1.7 0.4 - Iris-setosa
Ponto 7: 4.6 3.4 1.4 0.3 - Iris-setosa
Ponto 8: 5.3 4.1 5 0.2 - Iris-setosa
Ponto 9: 4.4 2.9 1.4 0.2 - Iris-setosa
Ponto 10: 4.9 3.1 1.5 0.1 - Iris-setosa
Ponto 11: 5.4 3.7 1.5 0.2 - Iris-setosa
Ponto 12: 4.8 3.4 1.6 0.2 - Iris-setosa
Ponto 13: 4.8 3 1.4 0.1 - Iris-setosa
Ponto 14: 4.3 3 1.1 0.1 - Iris-setosa
Ponto 15: 5.8 4 1.2 0.2 - Iris-setosa
Ponto 16: 5.7 4.4 1.5 0.4 - Iris-setosa
Ponto 17: 5.4 3.9 1.3 0.4 - Iris-setosa
Ponto 18: 5.1 3.5 1.4 0.3 - Iris-setosa
Ponto 19: 5.7 3.8 1.7 0.3 - Iris-setosa
Ponto 20: 5.1 3.8 1.5 0.3 - Iris-setosa
Ponto 21: 5.4 3.4 1.7 0.2 - Iris-setosa
Ponto 22: 5.1 3.7 1.5 0.4 - Iris-setosa
Ponto 23: 4.6 3.6 1 0.2 - Iris-setosa
Ponto 24: 5.1 3.3 1.7 0.5 - Iris-setosa
Ponto 25: 4.8 3.4 1.9 0.2 - Iris-setosa
Ponto 26: 5.3 1.6 0.2 - Iris-setosa
Ponto 27: 5.3 1.4 1.6 0.4 - Iris-setosa
Ponto 28: 5.2 3.5 1.5 0.2 - Iris-setosa
Ponto 29: 5.2 3.4 1.4 0.2 - Iris-setosa
Ponto 30: 4.7 3.2 1.6 0.2 - Iris-setosa
Ponto 31: 4.8 3.1 1.6 0.2 - Iris-setosa
Ponto 32: 5.4 3.4 1.5 0.4 - Iris-setosa
Ponto 33: 5.2 4.1 1.5 0.1 - Iris-setosa
Ponto 34: 5.5 4.2 1.4 0.2 - Iris-setosa
Ponto 35: 4.9 3.1 1.5 0.1 - Iris-setosa
Ponto 36: 5.3 2.2 1.2 0.2 - Iris-setosa
Ponto 37: 5.5 3.5 1.3 0.2 - Iris-setosa
Ponto 38: 4.9 3.1 1.5 0.1 - Iris-setosa
Ponto 39: 4.4 3 1.3 0.2 - Iris-setosa
Ponto 40: 5.1 3.4 1.5 0.2 - Iris-setosa
Ponto 41: 5.3 5 1.3 0.3 - Iris-setosa
Ponto 42: 4.5 2.3 1.3 0.3 - Iris-setosa
Ponto 43: 4.4 3.2 1.3 0.2 - Iris-setosa
Ponto 44: 5.3 5 1.6 0.6 - Iris-setosa
Ponto 45: 5.1 3.8 1.9 0.4 - Iris-setosa
Ponto 46: 4.8 3 1.4 0.3 - Iris-setosa
Ponto 47: 5.1 3.8 1.6 0.2 - Iris-setosa
Ponto 48: 4.6 3.2 1.4 0.2 - Iris-setosa
Ponto 49: 5.3 3.7 1.5 0.2 - Iris-setosa
Ponto 50: 5.3 3 1.4 0.2 - Iris-setosa
Atributos do Cluster: 5.006 3.418 1.464 0.244

Cluster 3

Ponto 134: 6.3 2.8 5.1 1.5 - Iris-virginica
Ponto 51: 7.3 2.4 7.1 4 - Iris-versicolor
Ponto 52: 6.4 3.2 4.5 1.5 - Iris-versicolor
Ponto 55: 6.5 2.8 4.6 1.5 - Iris-versicolor
Ponto 57: 6.3 3.3 4.7 1.6 - Iris-versicolor
Ponto 59: 6.6 2.9 4.6 1.3 - Iris-versicolor
Ponto 64: 6.1 2.9 4.7 1.4 - Iris-versicolor
Ponto 66: 6.7 3.1 4.4 1.4 - Iris-versicolor
Ponto 67: 5.6 3 4.5 1.5 - Iris-versicolor
Ponto 69: 6.2 2.2 4.5 1.5 - Iris-versicolor
Ponto 71: 5.9 3.2 4.8 1.8 - Iris-versicolor
Ponto 73: 6.3 2.5 4.9 1.5 - Iris-versicolor
Ponto 74: 6.1 2.8 4.7 1.2 - Iris-versicolor
Ponto 75: 6.4 2.9 4.3 1.3 - Iris-versicolor
Ponto 76: 6.6 3 4.4 1.4 - Iris-versicolor
Ponto 77: 6.8 2.8 4.8 1.4 - Iris-versicolor
Ponto 79: 6.2 4.5 1.5 - Iris-versicolor
Ponto 84: 6.2 7.5 1.6 - Iris-versicolor
Ponto 86: 6.3 4.5 1.6 - Iris-versicolor
Ponto 87: 6.7 3.1 4.7 1.5 - Iris-versicolor
Ponto 88: 6.3 2.3 4.4 1.3 - Iris-versicolor
Ponto 92: 6.1 3 4.6 1.4 - Iris-versicolor
Ponto 98: 6.2 2.9 4.3 1.3 - Iris-versicolor
Ponto 102: 5.8 2.7 5.1 1.9 - Iris-virginica
Ponto 114: 5.7 2.5 5 2 - Iris-virginica
Ponto 115: 5.8 2.8 5.1 2.4 - Iris-virginica
Ponto 120: 6.2 2.5 1.5 - Iris-virginica
Ponto 122: 5.6 2.8 4.9 2 - Iris-virginica
Ponto 124: 6.3 2.7 4.9 1.8 - Iris-virginica
Ponto 127: 6.2 2.8 4.8 1.8 - Iris-virginica
Ponto 128: 6.1 3 4.9 1.8 - Iris-virginica
Ponto 139: 6.3 4.8 1.8 - Iris-virginica
Ponto 143: 5.8 2.7 5.1 1.9 - Iris-virginica
Ponto 147: 6.3 2.5 5 1.9 - Iris-virginica
Ponto 150: 5.9 3 5.1 1.8 - Iris-virginica
Ponto 54: 5.5 2.3 4 1.3 - Iris-versicolor
Ponto 56: 5.7 2.8 4.5 1.3 - Iris-versicolor
Ponto 60: 5.2 2.7 3.9 1.4 - Iris-versicolor
Ponto 62: 5.9 3 4.2 1.5 - Iris-versicolor
Ponto 63: 6.2 4 1 - Iris-versicolor
Ponto 68: 5.8 2.7 4.1 1 - Iris-versicolor
Ponto 70: 5.6 2.5 3.9 1.1 - Iris-versicolor
Ponto 72: 6.1 2.8 4 1.3 - Iris-versicolor
Ponto 81: 5.5 2.4 3.8 1.1 - Iris-versicolor
Ponto 83: 5.8 2.7 3.9 1.2 - Iris-versicolor
Ponto 85: 5.4 3 4.5 1.5 - Iris-versicolor
Ponto 89: 5.6 3 4.1 1.3 - Iris-versicolor
Ponto 90: 5.5 2.5 4 1.3 - Iris-versicolor
Ponto 91: 5.5 2.6 4.4 1.2 - Iris-versicolor
Ponto 93: 5.8 2.6 4 1.2 - Iris-versicolor
Ponto 95: 5.6 2.7 4.2 1.3 - Iris-versicolor
Ponto 96: 5.7 3 4.2 1.2 - Iris-versicolor
Ponto 97: 5.7 2.9 4.2 1.3 - Iris-versicolor

Ponto 97: 5.7 2.9 4.2 1.3 - Iris-versicolor
Ponto 100: 5.7 2.8 4.1 1.3 - Iris-versicolor
Ponto 107: 4.9 2.5 4.5 1.7 - Iris-virginica
Ponto 61: 5 2 3.5 1 - Iris-versicolor
Ponto 65: 5.6 2.9 3.6 1.3 - Iris-versicolor
Ponto 80: 5.7 2.6 3.5 1 - Iris-versicolor
Ponto 82: 5.5 2.4 3.7 1 - Iris-versicolor
Ponto 94: 5 2.3 3.3 1 - Iris-versicolor
Ponto 58: 4.9 2.4 3.3 1 - Iris-versicolor
Ponto 99: 5.1 2.5 3 1.1 - Iris-versicolor
Atributos do Cluster: 5.90161 2.74839 4.39355 1.43387

Tempo passado para o processo de agrupamento: 0.004

Porcentagem de pontos incorretos no cluster: 1
5.3%

Porcentagem de pontos incorretos no cluster: 2
0%

Porcentagem de pontos incorretos no cluster: 3
23%

Porcentagem total de pontos agrupados incorretamente: 11%
Tempo passado para o processo de verificacao: 0.008

Considerações Finais