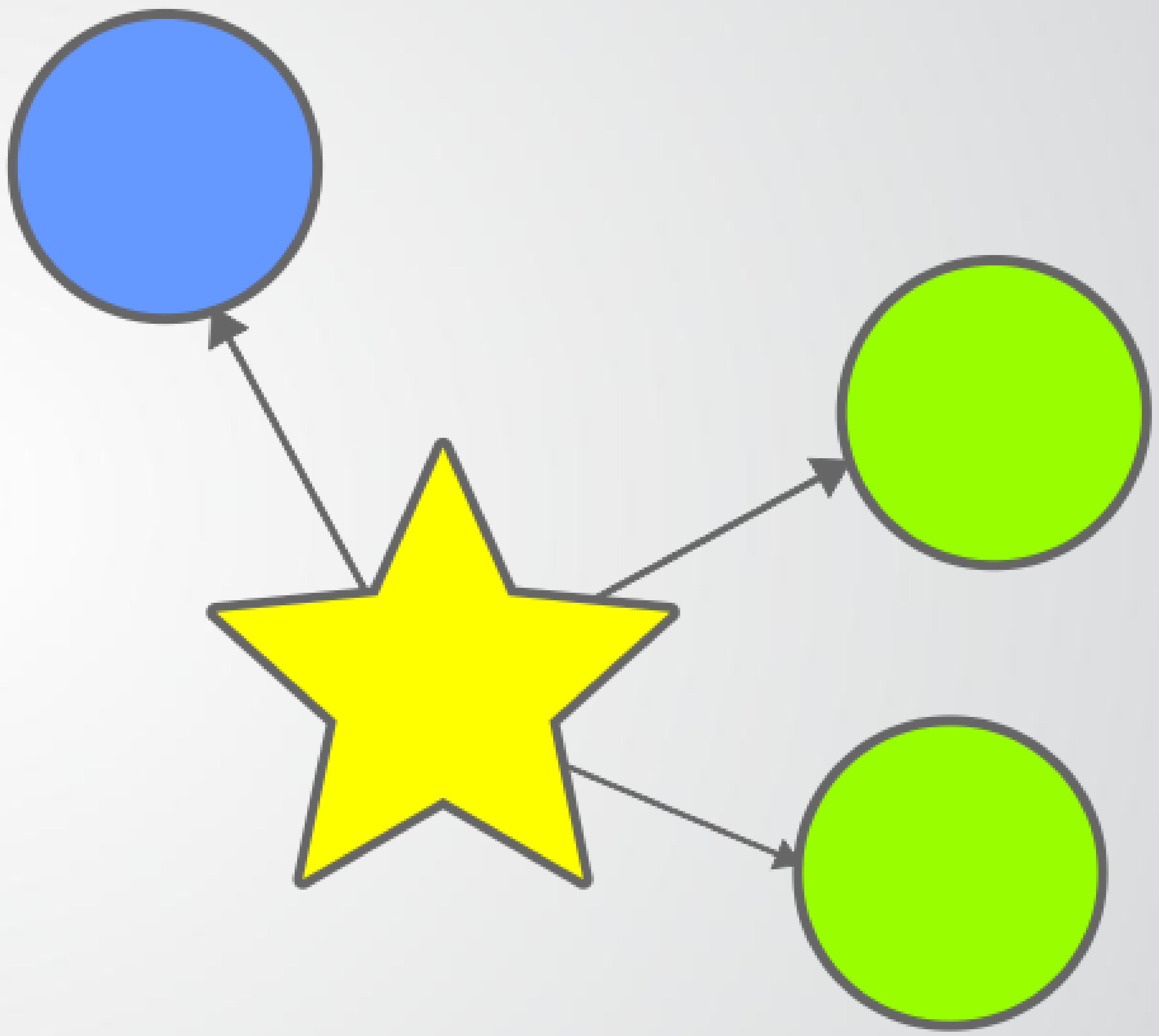
/ CANALSANDECO

K NEAREST NEIGHBOR



K VIZINHOS MAIS PRÓXIMOS

/CANALSANDECO

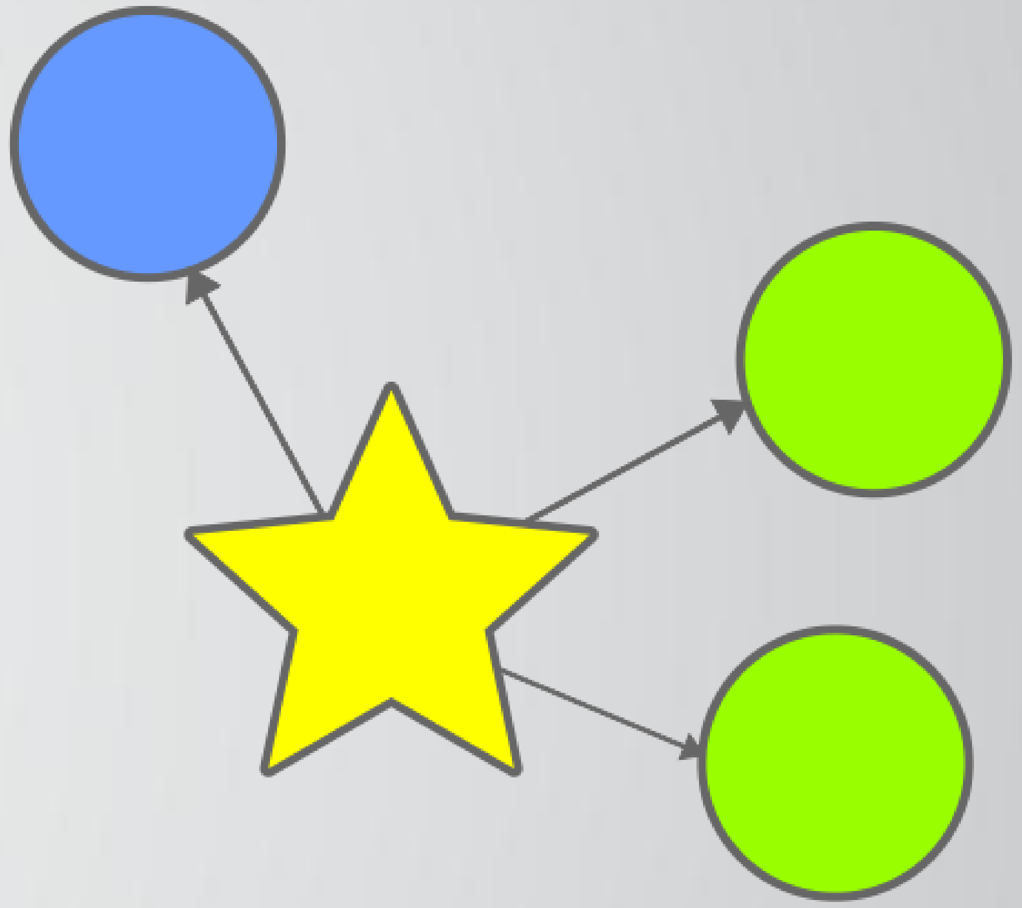
KNN

MÉTODO BASEADO EM DISTÂNCIAS

É UMA TÉCNICA QUE CONSIDERA

A PROXIMIDADE ENTRE DADOS

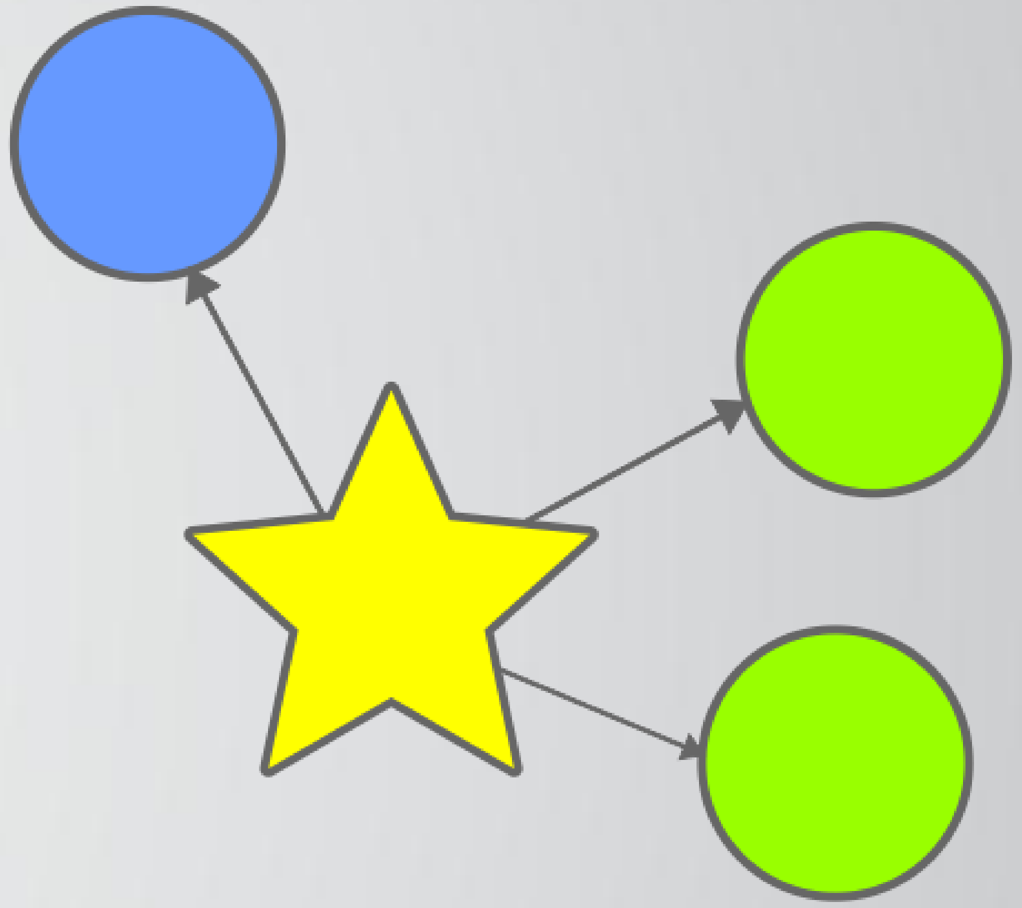
NA REALIZAÇÕES DE PREDIÇÕES



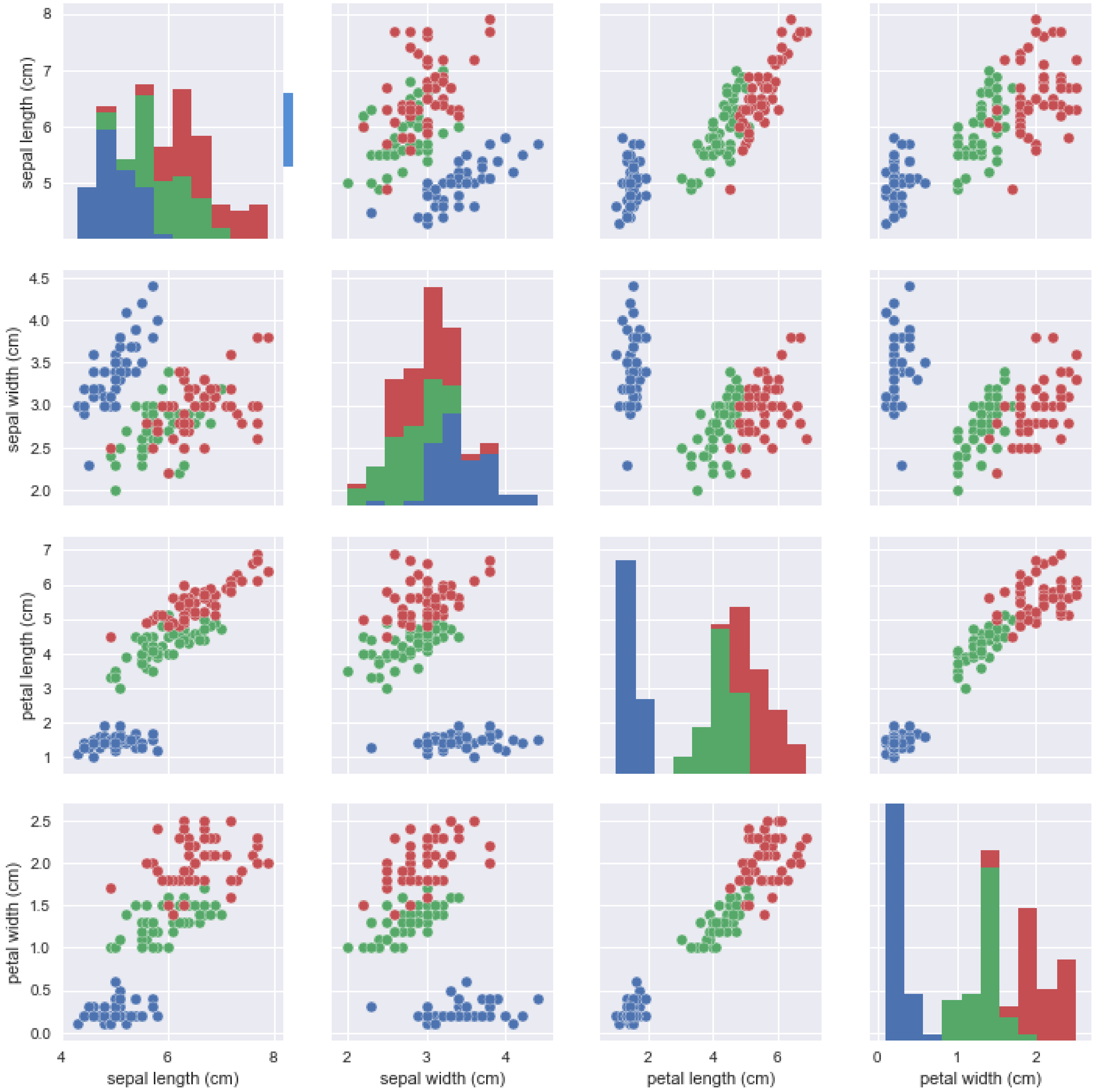
KNN

HIPÓTESE

DADOS SIMILARES TENDEM A ESTAR CONCENTRADOS NA MESMA REGIÃO NO ESPAÇO DE ***DISPERSÃO*** DOS DADOS



IRIS FLOWER DISPERSION

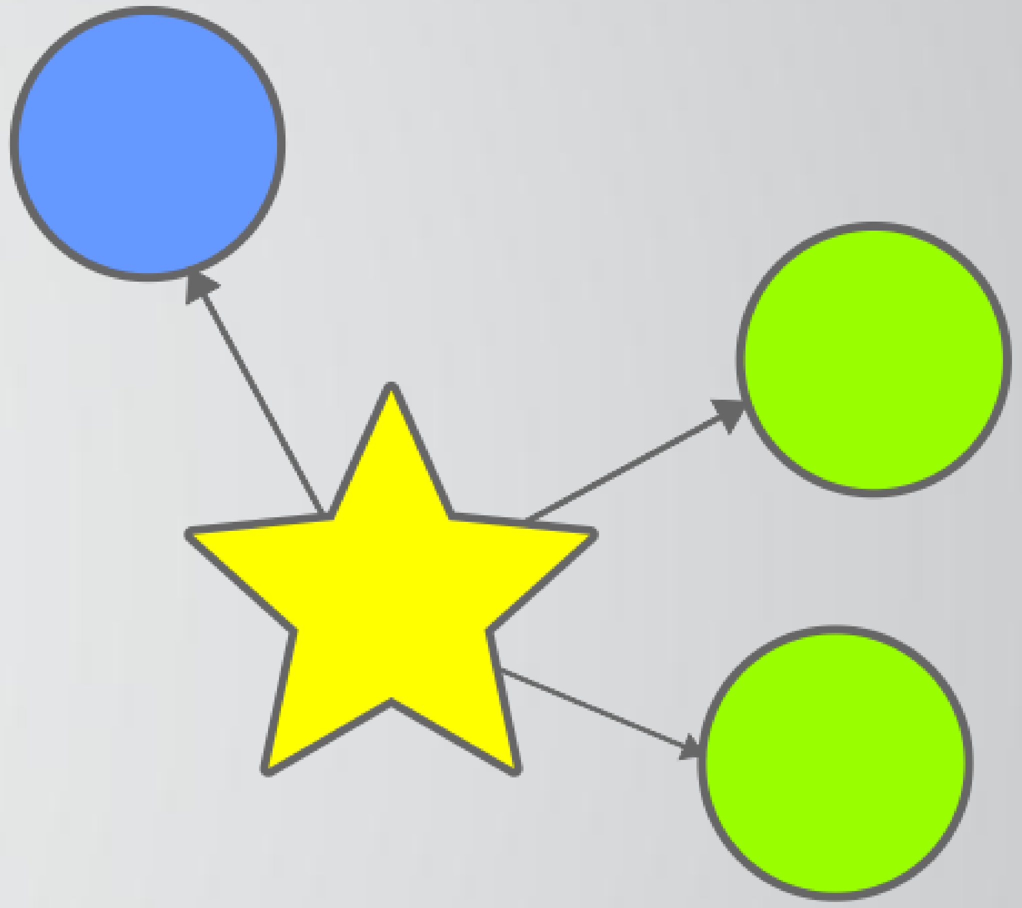


CLASSIFICAÇÃO

KNN

INTUIÇÃO

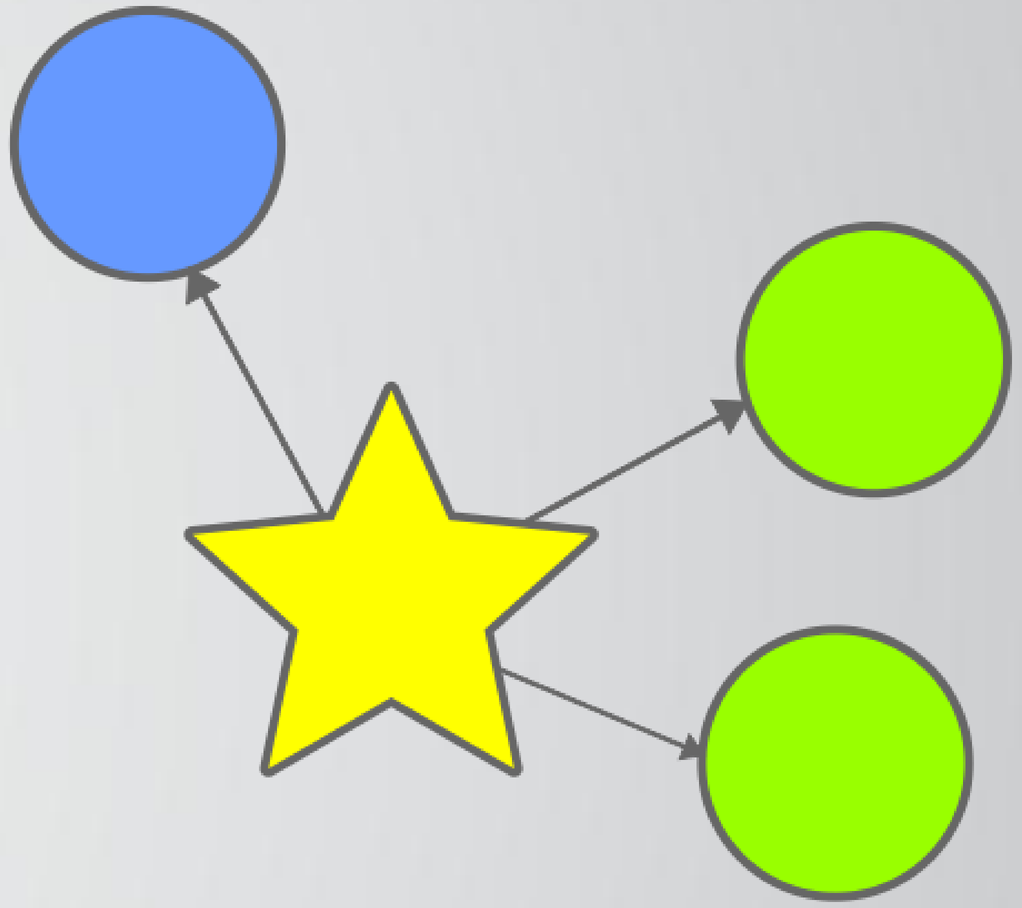
OBJETOS RELACIONADOS AO MESMO CONCEITO SÃO **SEMELHANTES** ENTRE SI



KNN

VANTAGEM

PODE SER USADO PARA CLASSIFICAÇÃO COMO PARA REGRESSÃO, SEM ALTERAÇÕES SIGNIFICATIVAS.



4 MEDIDAS DE DISTÂNCIAS

KNN

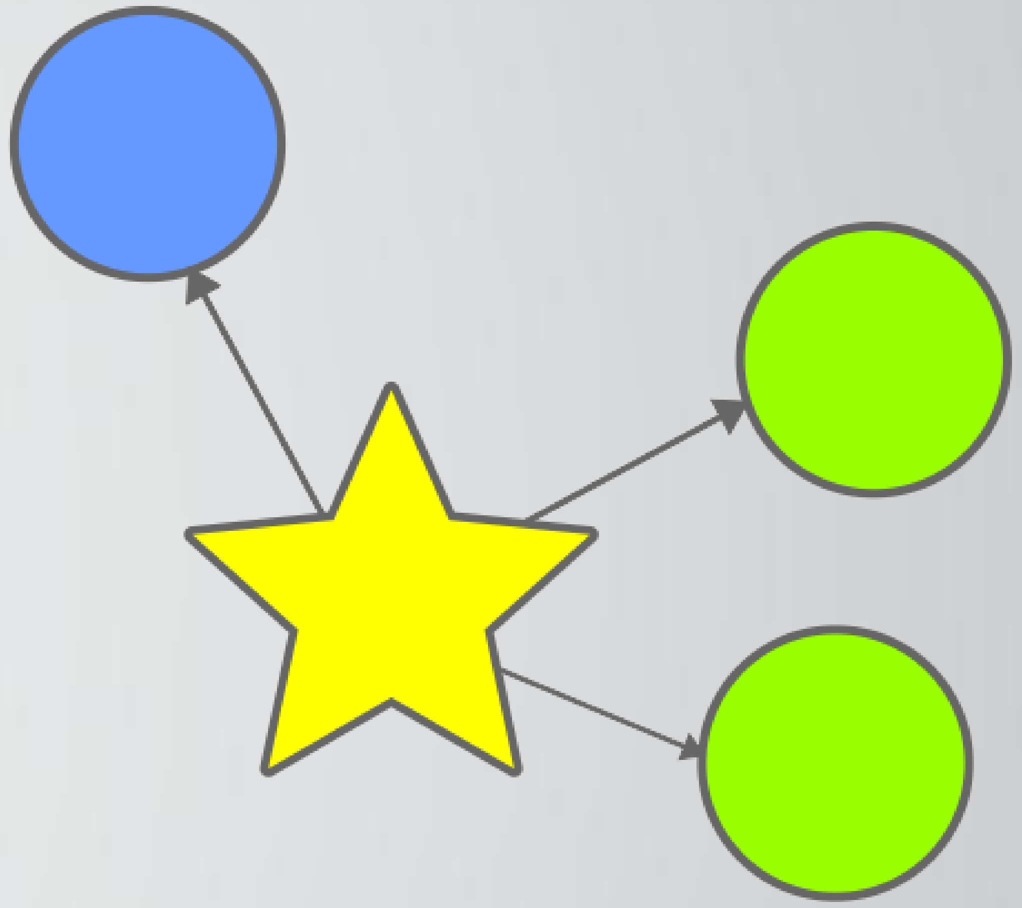
DISTÂNCIAS

DISTÂNCIA EUCLIDIANA

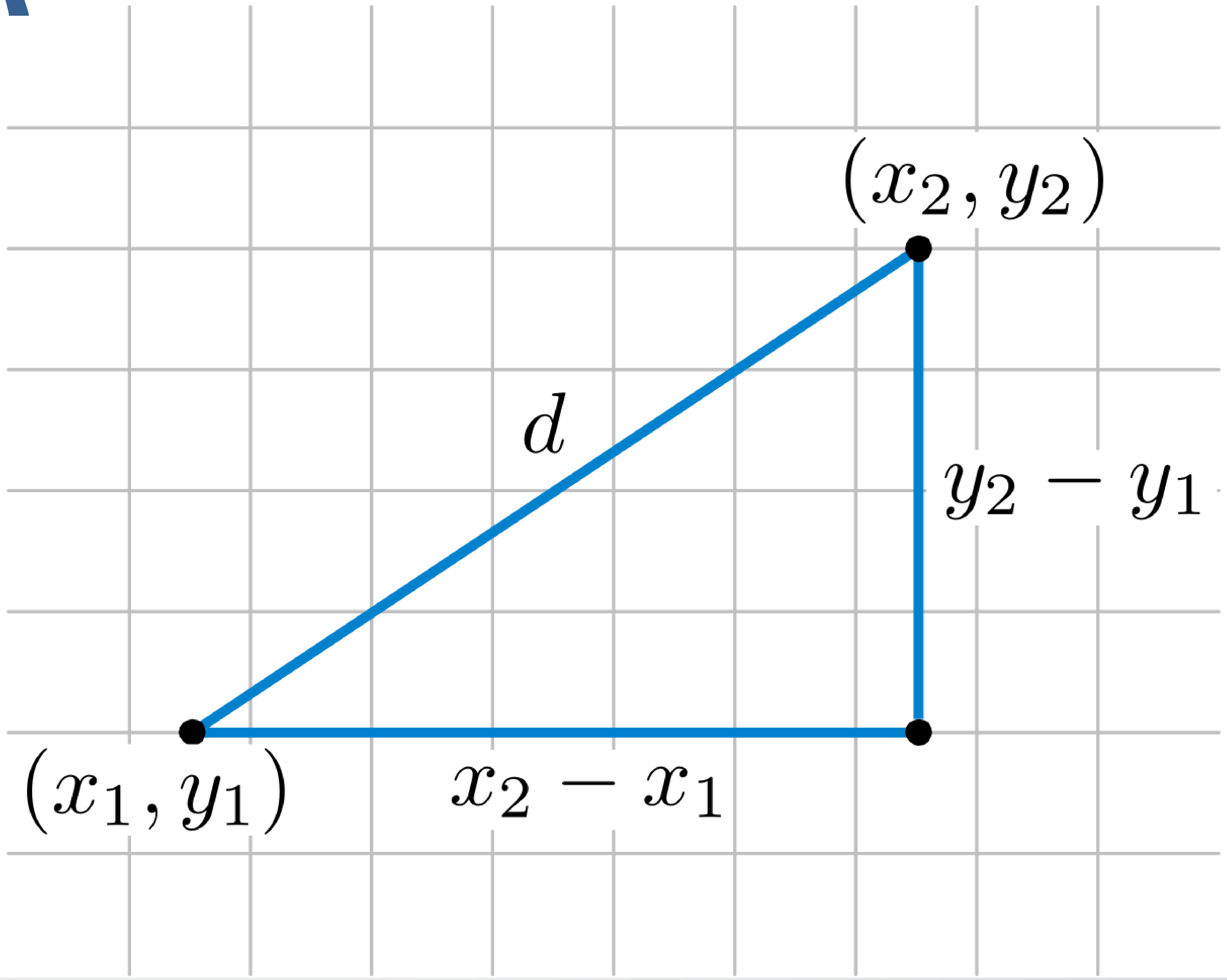
DISTÂNCIA DE MANHATTAN

DISTÂNCIA MINKOWSKI

DISTÂNCIA JACCARD



DISTÂNCIA EUCLIDIANA



Sdeco

DISTÂNCIA EUCLIDIANA

The **Euclidean distance** between points **p** and **q** is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).

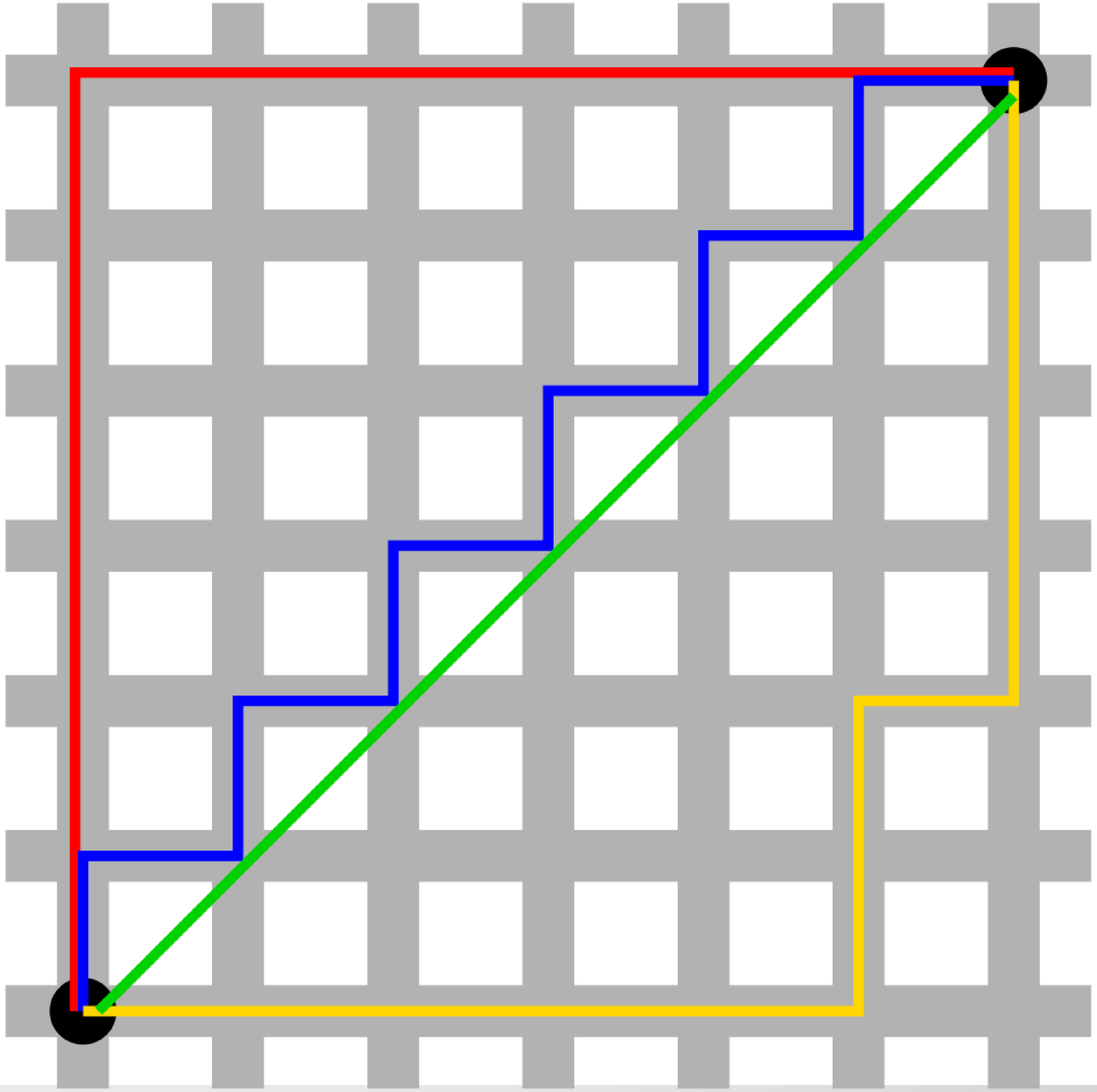
In Cartesian coordinates, if **p** = (p_1, p_2,..., p_n) and **q** = (q_1, q_2,..., q_n) are two points in Euclidean *n*-space, then the distance (d) from **p** to **q**, or from **q** to **p** is given by the Pythagorean formula:

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

DISTÂNCIA MANHATTAN

GEOMETRIA DO TAXI



DISTÂNCIA MANHATTAN

The taxicab distance, d_1, between two vectors $\mathbf{p}, \mathbf{q}$ in an n-dimensional real vector space with fixed Cartesian coordinate system, is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. More formally,

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|,$$

where $(\mathbf{p}, \mathbf{q})$ are vectors

$$\mathbf{p} = (p_1, p_2, \dots, p_n) \text{ and } \mathbf{q} = (q_1, q_2, \dots, q_n)$$

For example, in the plane, the taxicab distance between (p_1, p_2) and (q_1, q_2) is $|p_1 - q_1| + |p_2 - q_2|$.

DISTÂNCIA MINKOWSKI

A distância de Minkowski é uma forma métrica generalizada de distância euclidiana e distância de Manhattan.

DISTÂNCIA MINKOWSKI

The Minkowski distance of order *p* between two points

$$X = (x_1, x_2, \ldots, x_n) \text{ and } Y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$$

is defined as:

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$$

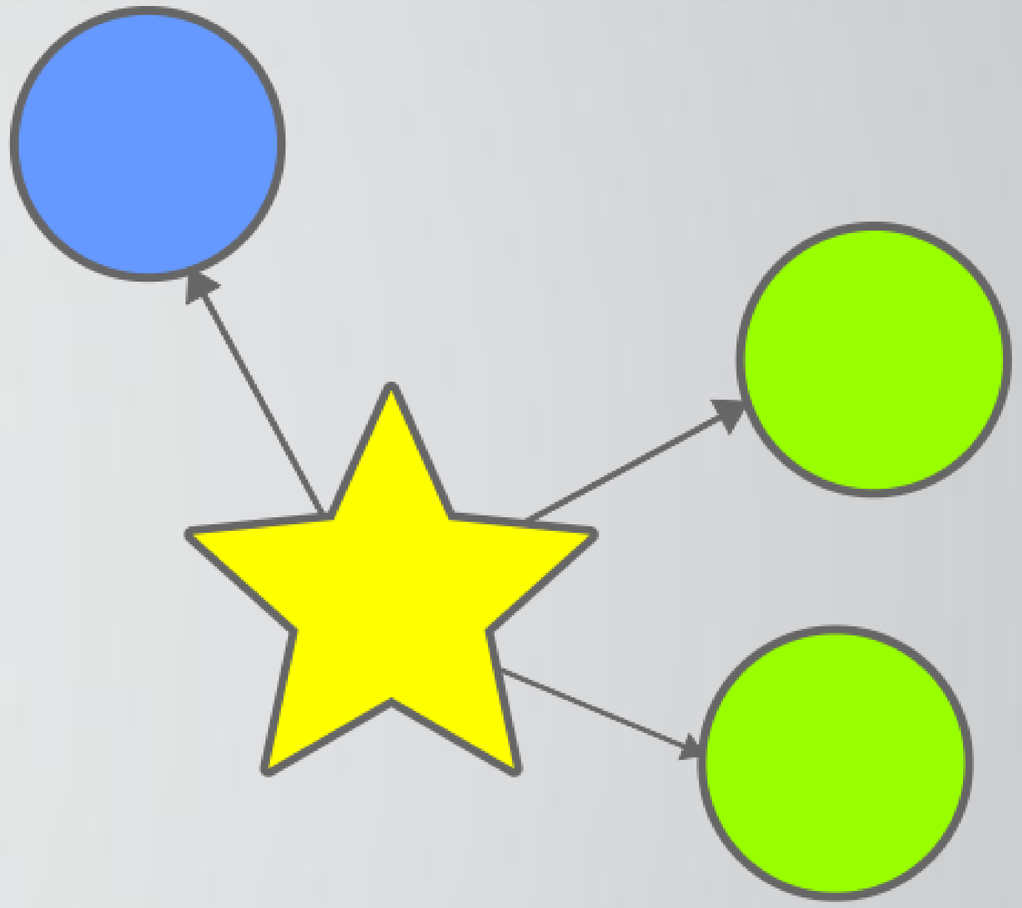
For $p \geq 1$, the Minkowski distance is a metric as a result of the Minkowski inequality. When $p < 1$, the distance between (0,0) and (1,1) is $2^{1/p} > 2$, but the point (0,1) is at a distance 1 from both of these points. Since this violates the triangle inequality, for $p < 1$ it is not a metric.

Minkowski distance is typically used with *p* being 1 or 2, which correspond to the Manhattan distance and the Euclidean distance, respectively.

KNN

PASSOS DO ALGORITMO

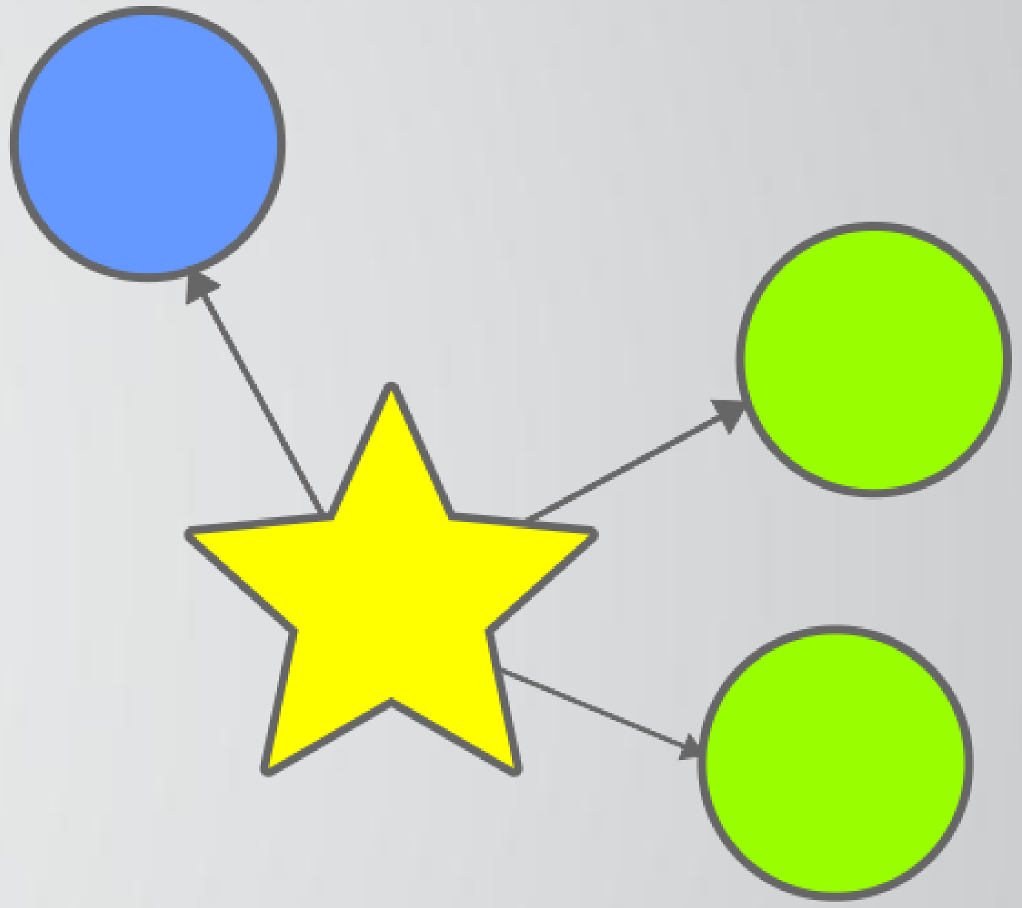
PASSOS PARA CLASSIFICAR UM NOVO EXEMPLO UM OBJETO ESTRELA ->



KNN

PASSO 1

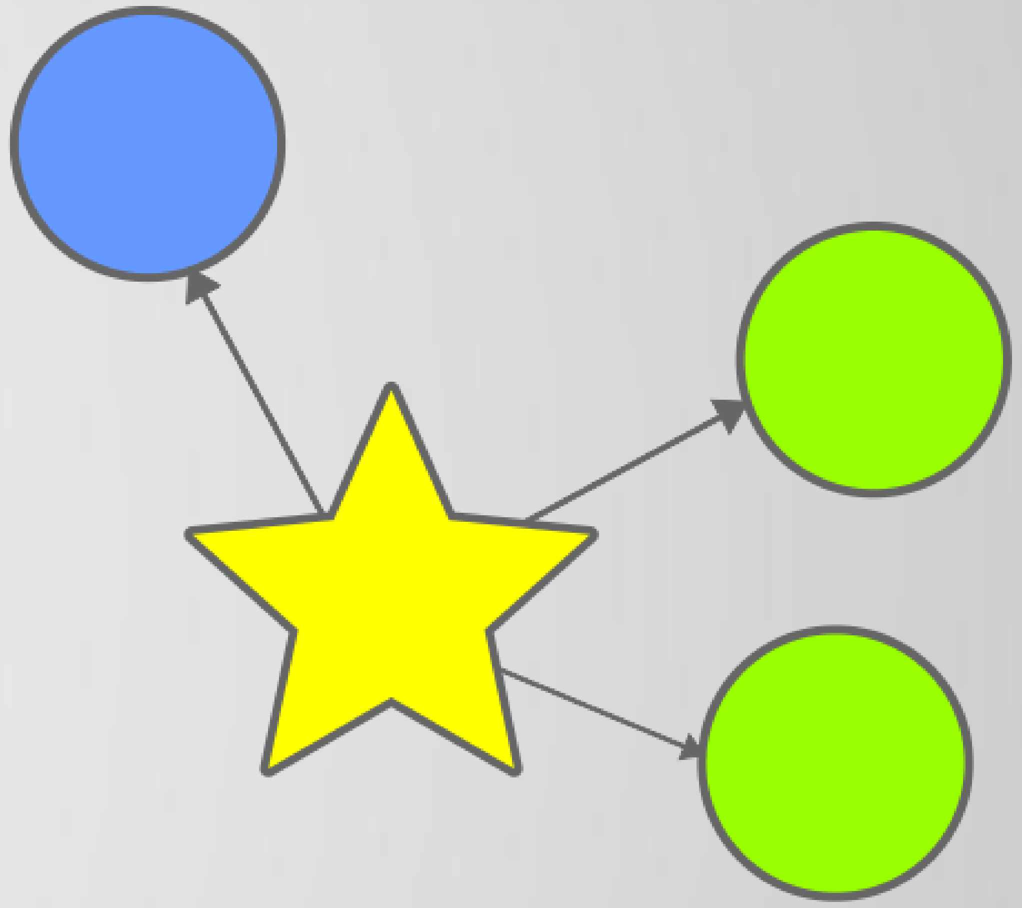
CALCULE A DISTÂNCIA EUCLIDIANA ENTRE O NOVO OBJETO E TODOS OS OBJETOS.



KNN

PASSO 2

ORDENE DE FORMA ASCENDENTE AS DISTÂNCIAS CALCULADAS.

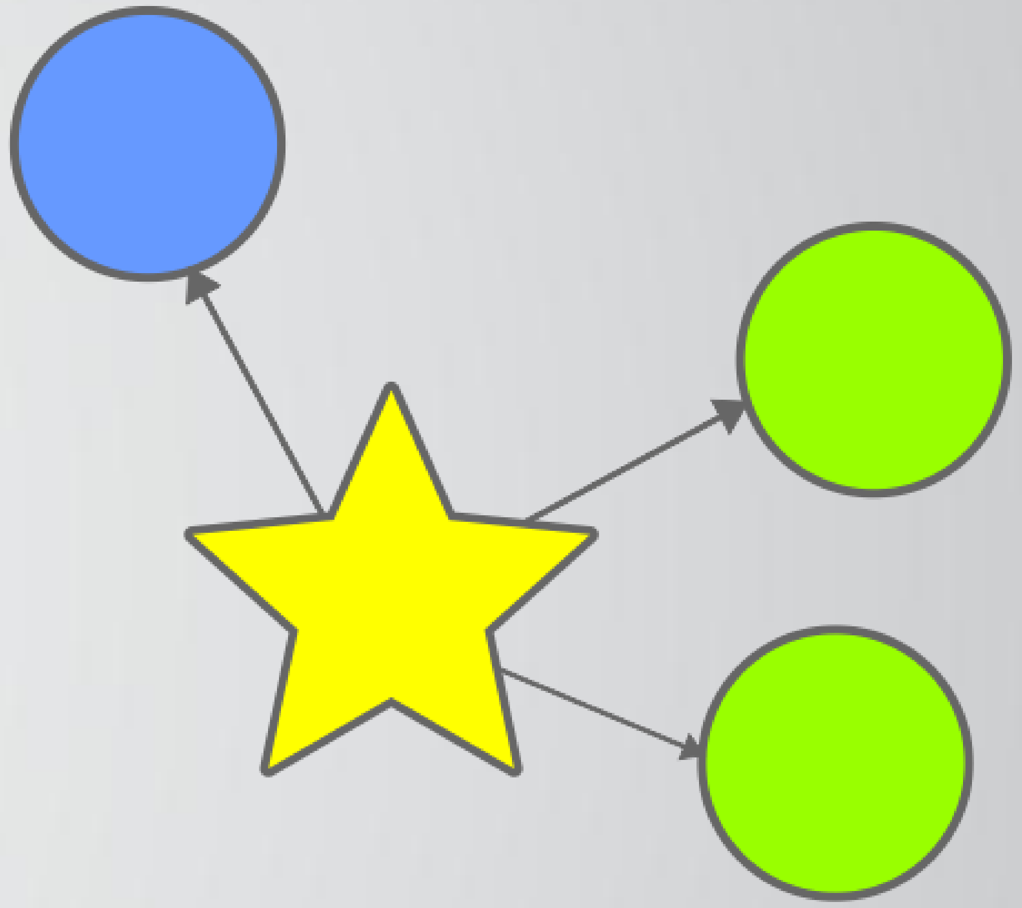


CLASSIFICAÇÃO

KNN

PASSO 3

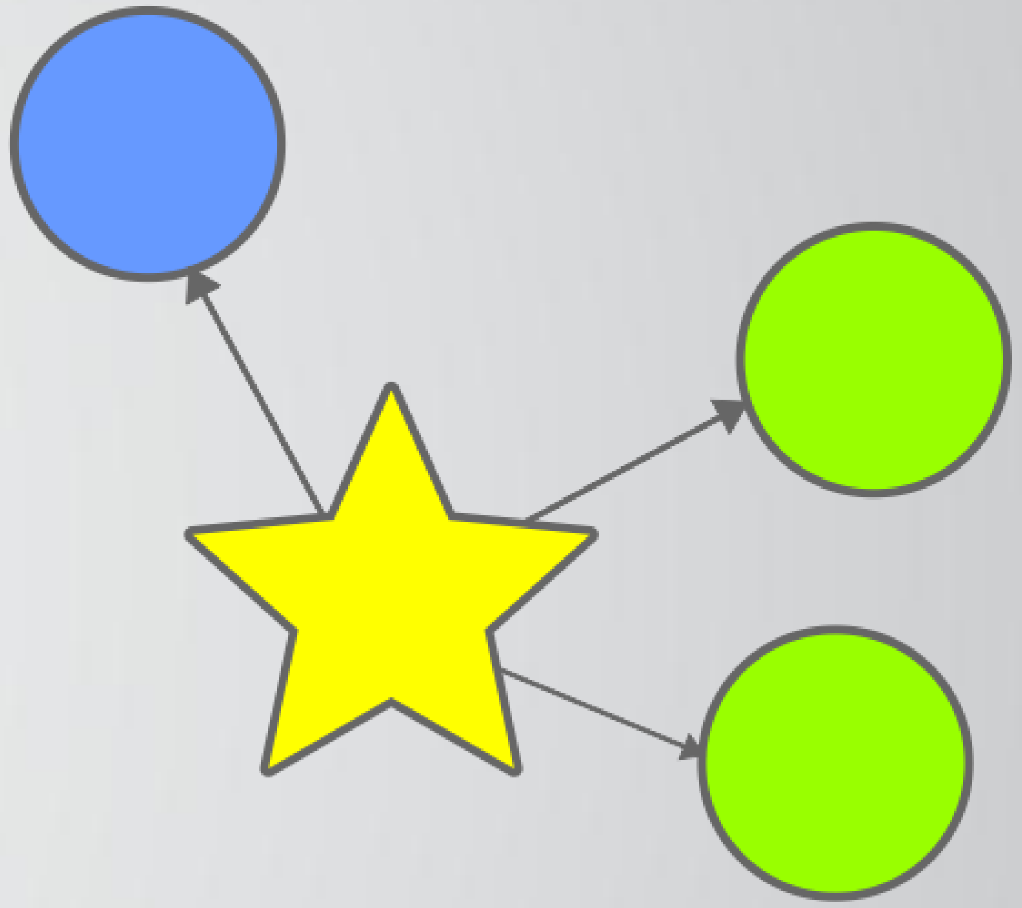
REALIZE A SOMATÓRIA DAS CLASSES ATÉ ‘K’ EXEMPLARES.



KNN

PASSO 4

ATRIBUA A CLASSE GANHADORA AO NOVO OBJETO



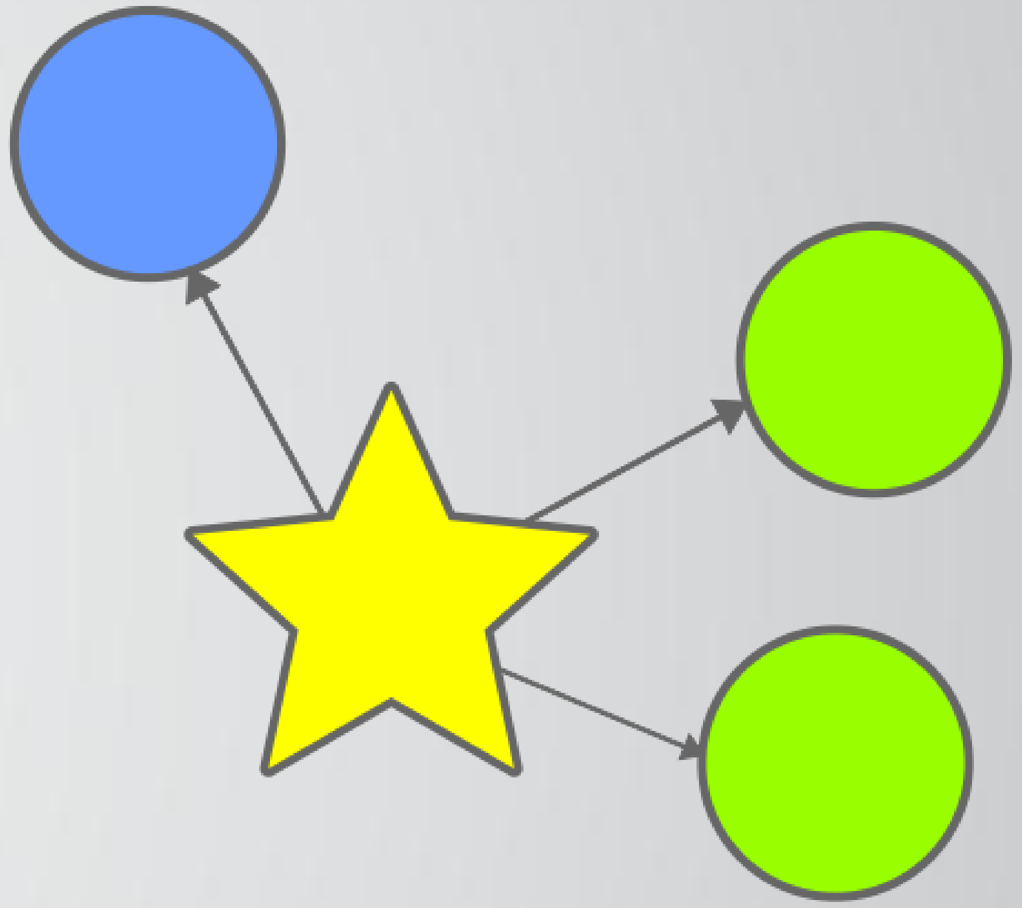
/CANALSANDECO

CLASSIFICAÇÃO

KNN

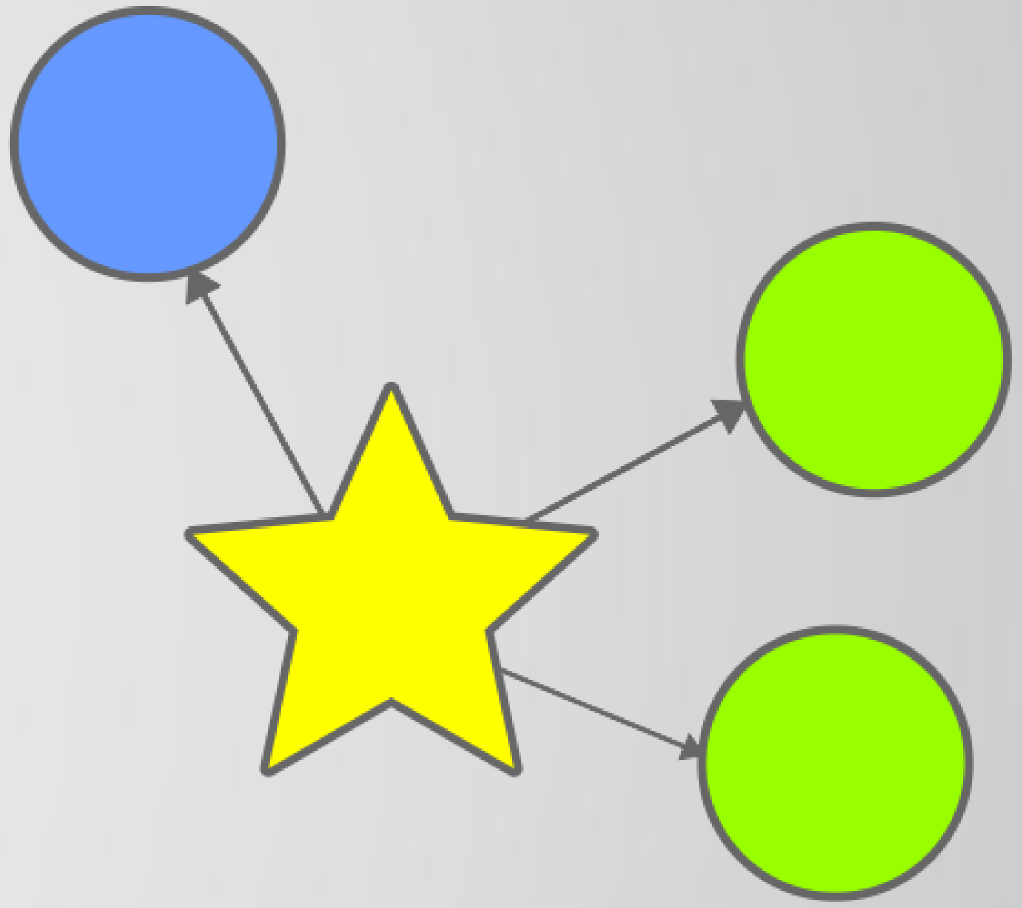
DICA

PARA EVITAR EMPATE ENTRE AS CLASSE ESCOLHA UM VALOR DE **'K'** IMPAR



KNN

PARA EVITAR EMPATE ENTRE AS CLASSE ESCOLHA UM VALOR DE **‘K’** IMPAR



SCIKIT LEARN

```
from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n_neighbors=3)

knn.fit(Features, classes)

Knn.predict(new_object)
```