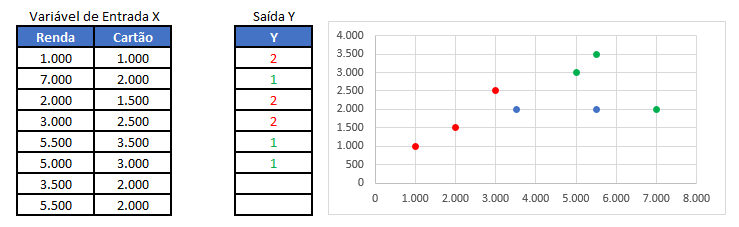
**KNN = K-vizinhos Mais Próximos**



2 por proximidade

1 por proximidade

* O “motor” de funcionamento do KNN está na descoberta dos objetos mais próximos.
* Para usá-lo precisamos:
  1. Definir um conjunto de treinamento */já rotuladas*
  2. Definir uma métrica de distância
  3. Definir o valor de K
  4. Normalizar os dados */mesma ordem (trazer as variáveis para o mesmo domínio e mesma significância)*

CLASSIFICANDO UM NOVO EXEMPLO:

1. Calcule a distância entre o novo objeto em relação aos objetos de treinamento
2. Identifique os K mais próximo
3. Identifique a classe dos K-Vizinhos mais próximos para determinar o rótulo do novo objeto

VANTAGENS

* Simples e fácil de implementar
* Apresenta resultados competitivos

DESVANTAGENS

* Caro para bases de treino “grandes”
* Sensível à variáveis irrelevantes ou a ruídos

\_\_\_\_\_\_\_\_\_\_\_

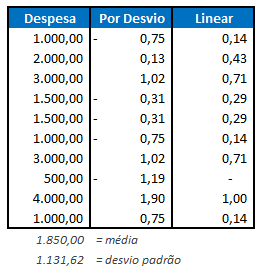
* Distância Euclidiana “d”: Seja “P” e “Q” objetos de classificação: */linha reta*
* City-Block: */soma das distâncias por blocos*
* Outras distâncias
  + Hamming
  + Mahalanobis

\_\_\_\_\_\_\_\_\_\_\_

* Normalização Linear [0,1]
* Normalização por Desvio Padrão considera a posição médio nos valores e os graus de dispersão à média: */útil quando o Mínimo e Máximo são desconhecidos.*

***Variância:*** *Medida da dispersão em relação à média*

*S =*

**

**LDA (Linear Discriminant Analisys) e QDA (Quadratic Discriminant Analisys)**

* Classificadores Alternativos ao Naive Bayes
* Considera as variáveis de forma multivariada
* Suposição de que os dados estão normalmente distribuídos

**Distribuição Normal Univariada**

f (x) =

**Distribuição Normal Multivariada**

f (x) =

Dada duas classes e em que a probabilidade de um objeto x se associado a uma das classes sejam iguais P() = P(), a superfície de separação entre as duas classes e é dada:

=

em que de dimensão Kx1 é o vetor média e de KxK é a matriz de covariâncias para a classe j.

Aplicando o logaritmo dos dois lados:

**Algoritmo QDA:**

1. Dado X e Y
2. Calcule e
3. Fim treinamento

Dado um novo objeto .

**D1 =**

**D2 =**

O agrupamento é feito conforme proximidade do valor de em relação a D1 ou D2.

**Algoritmo LDA:**

Calculamos uma única matriz de covariâncias para todos os objetos.

**D1 =**

