

# Learning Vector Quantization – LVQ

Prof. Dr. Ernane Costa



# Contextualizando...



## Um problema de classificação

---

Título: Zoo database

Fonte da informação

- Creator: Richard Forsyth
- Site: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/zoo/>

Informação Relevante:

- Conjunto de dados simples contendo 17 atributos descritivos.
- O atributo "type" aparece como identificador da classe do atributo.
- Assunto: Qual animal é de qual tipo.

Valores de atributos *Missing*: nenhum

Distribuição de classes: na tabela

# Um problema de classificação



Classe	#	Conjunto de Animais
1	41	aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf
2	20	chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren
3	5	pitviper, seasnake, slowworm, tortoise, tuatara
4	13	bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna
5	4	frog, frog, newt, toad
6	8	flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
7	10	clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm

# Um problema de classificação



Número de instâncias (dados): 101

Número de atributos: 18 (nome do animal (único), 15 atributos booleanos, 2 numéricos)

Informação sobre os atributos: (nome e domínio)

1. animal name: único para cada dado
2. hair: Boolean
3. feathers: Boolean
4. eggs: Boolean
5. milk: Boolean
6. airborne: Boolean
7. aquatic: Boolean
8. predator: Boolean
9. toothed: Boolean
10. backbone: Boolean
11. breathes: Boolean
12. venomous: Boolean
13. fins: Boolean
14. legs: Numeric (set of values: {0,2,4,5,6,8})
15. tail: Boolean
16. domestic: Boolean
17. catsize: Boolean
18. type: Numeric (integer values in range [1,7]) - informação de classe

# Um problema de classificação



## Uma parte do conjunto de dados

- Para treinamento

aardvark, 1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1

← Dado original

Identificador	→	antelope,	1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,	
		bass,	0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,	4
		bear,	1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,	
		boar,	1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,	

← Informação de classe

- Para teste

↑  
Atributos descritivos

↓  
1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1  
1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1  
0,0,1,0,0,1,0,1,1,0,0,1,0,1,1,0

# LVQ –Aspectos Gerais



Rede Neural Artificial proposta por Teuvo Kohonen.

Baseada em competição (aprendizado competitivo e supervisionado)

É um método de classificação de padrões no qual cada unidade de saída representa uma classe em particular.

- Várias unidades de saída deveriam ser usadas para cada classe.

O vetor de pesos para uma unidade de saída é freqüentemente chamado de vetor *codebook* para a classe que a unidade de saída representa.

Durante o treinamento, as unidades de saída são posicionadas de forma a aproximar a superfície de decisão do classificador.

# LVQ –Aspectos Gerais



- Entrada para o treinamento:
  - Um conjunto de padrões de treinamento com classificação conhecida.
  
- Após o treinamento a rede LVQ é capaz de classificar um vetor de entrada por meio da associação dele a uma unidade de saída que representa uma classe.
  - Tal unidade de saída tem seu vetor de pesos (vetor codebook) próximo (vetorialmente) ao vetor de entrada.

# LVQ – Algoritmo de treinamento



## ▣ Nomenclatura:

- $\mathbf{x}$  – vetor de treinamento  $(x_1, \dots, x_i, \dots, x_n)$
- $T$  – classe correta para o vetor de treinamento
- $\mathbf{w}_j$  – vetor peso da  $j$ -ésima unidade de saída  $(w_{1,j}, \dots, w_{i,j}, \dots, w_{n,j})$
- $C_j$  – classe representada pela  $j$ -ésima unidade de saída
- $\| \mathbf{x} - \mathbf{w}_j \|$  - distância Euclidiana entre o vetor de entrada e (vetor de pesos para) o  $j$ -ésimo vetor de saída.
- $J$  - uma unidade de saída



# LVQ – Algoritmo de treinamento



- ▣ Passo 0: Inicialize os vetores de peso;  
inicialize a taxa de aprendizado,  $\alpha(0)$ .

## Inicialização dos vetores de peso:

- Usar os primeiros  $m$  vetores como vetores de peso;
- Associação randômica;
- Usar o K-means (algoritmo clássico de clusterização) ou uma rede auto-organizável.

## O que é a taxa de aprendizado?

- Fator que determina a velocidade do aprendizado;
- Passo do aprendizado;
- Regulador de convergência.

# LVQ – Algoritmo de treinamento



Passo 1:

**Enquanto** a condição de parada é falsa, execute os passos 2-6.

- Passo 2:

**Para cada** vetor de entrada de treinamento  $\mathbf{x}$ , execute os passos 3-4.

- Passo 3:

Encontre  $\mathbf{J}$  tal que  $\| \mathbf{x} - \mathbf{w}_{\mathbf{J}} \|$  seja mínima.

# LVQ – Algoritmo de treinamento



- Passo 4:

Altere  $w_j$  como segue:

**Se  $T = C_j$ , então**

$$w_j(\text{new}) = w_j(\text{old}) + \alpha[x - w_j(\text{old})];$$

**Se  $T \neq C_j$ , então**

$$w_j(\text{new}) = w_j(\text{old}) - \alpha[x - w_j(\text{old})];$$

- Passo 5:

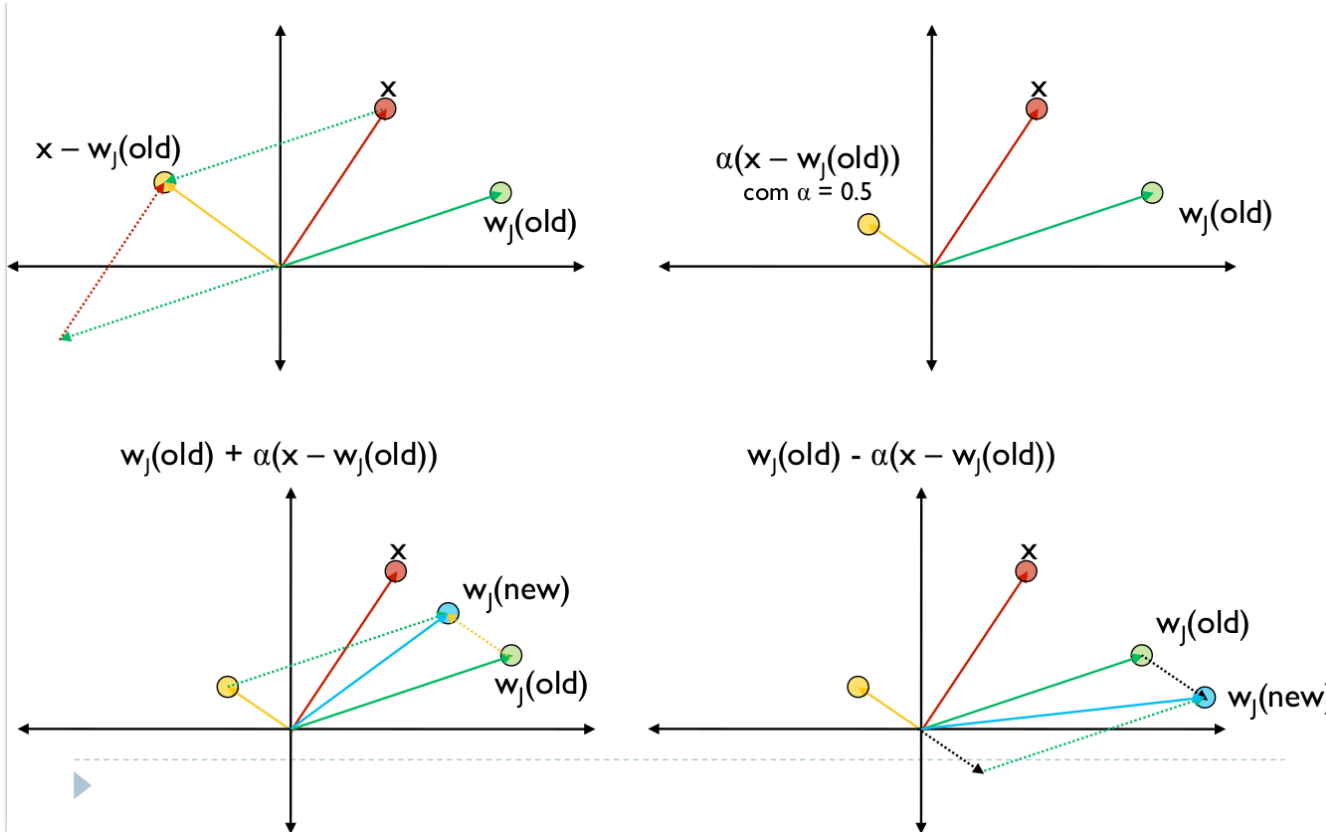
Reduza a taxa de aprendizado.

- Passo 6:

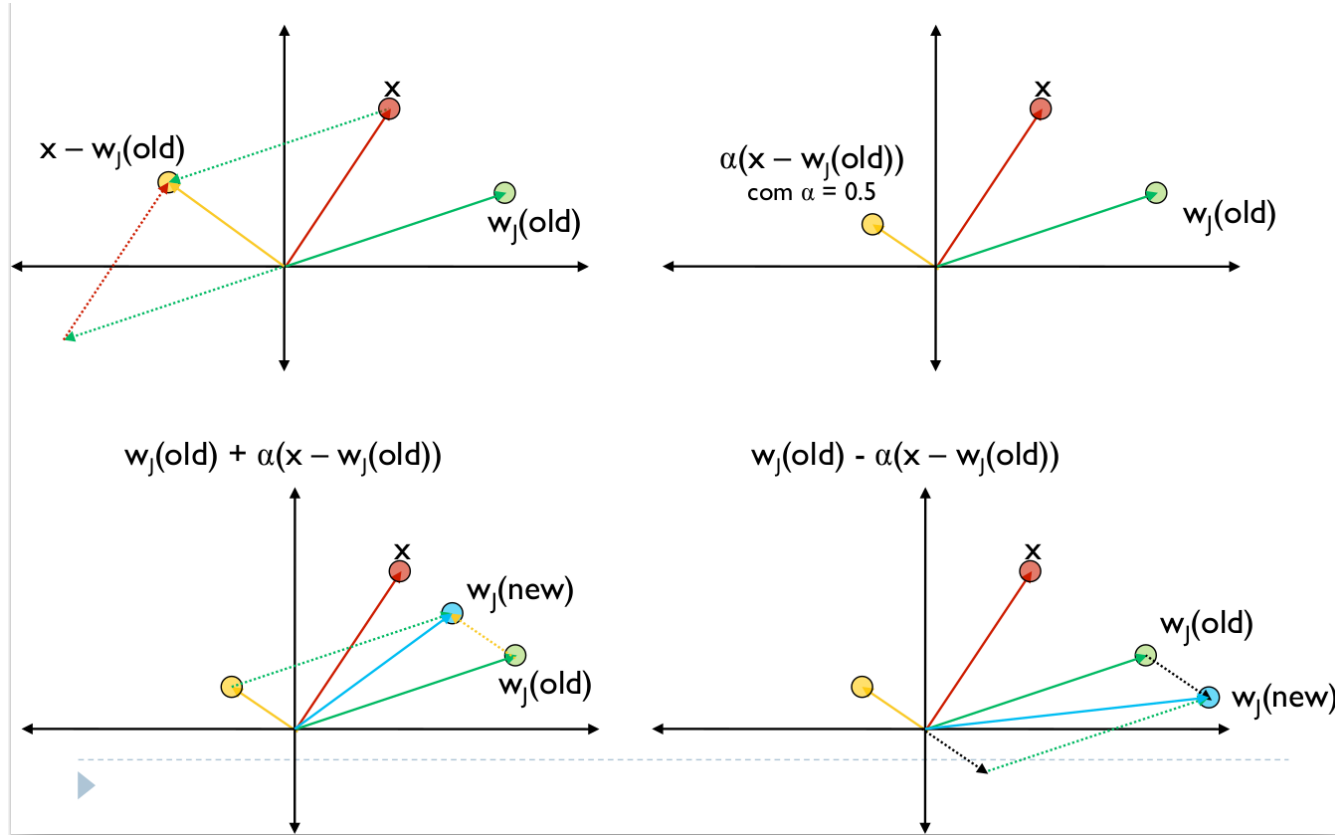
Teste a condição de parada

A condição deve especificar um número fixo de iterações (i.e., execução do Passo 1) ou um valor mínimo para a taxa de aprendizado.

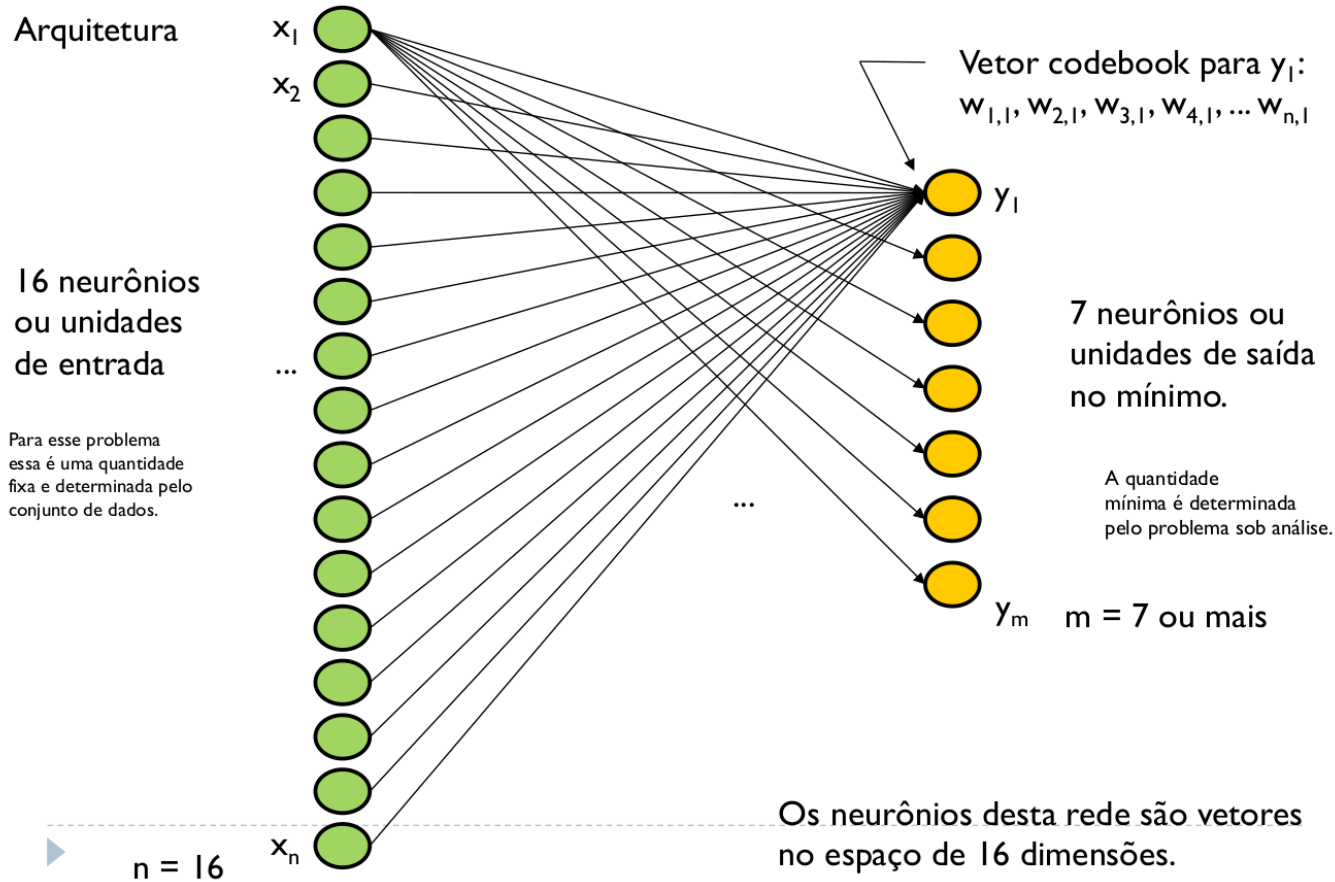
# LVQ – Analisando a alteração de peso



# LVQ – Analisando a alteração de peso



# LVQ – Apresentando a rede neural artificial

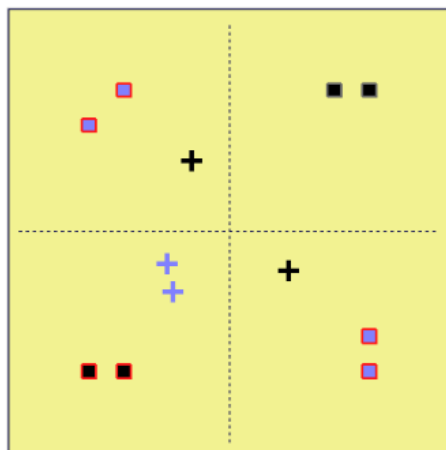


# LVQ – Entendendo o funcionamento



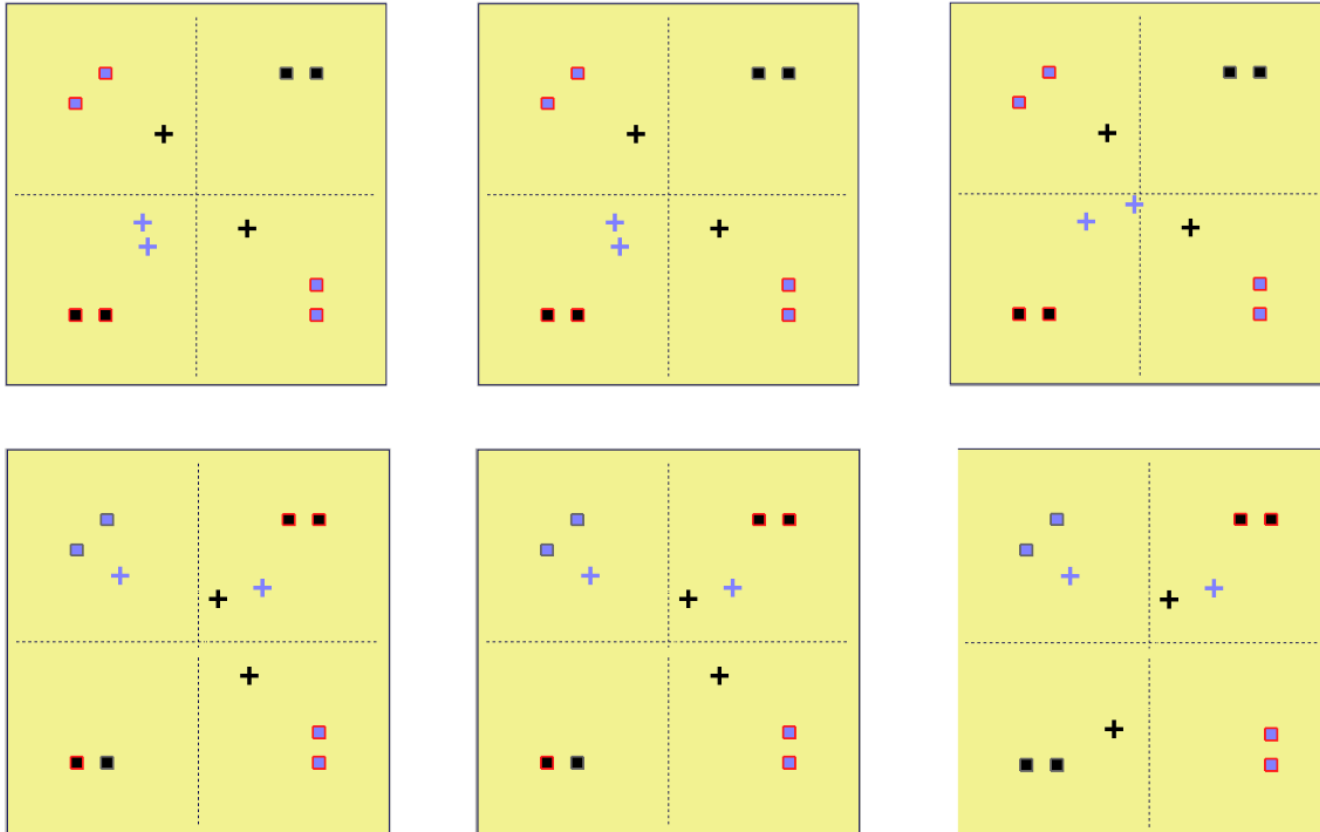
Motivação: encontrar a unidade de saída que está mais próxima do vetor de entrada.

- Se  $x$  e  $y$  pertencem à mesma classe, então mova os pesos em direção ao vetor de entrada;
- Se  $x$  e  $y$  pertencem a classes diferentes, então mova os pesos em direção oposta ao vetor de entrada;



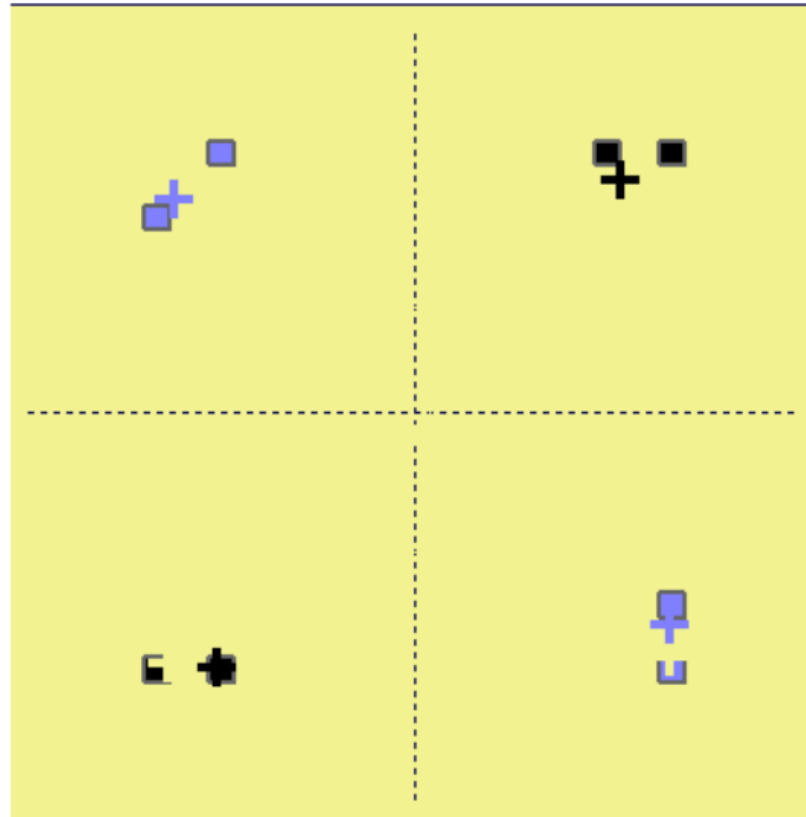
Quadrados: vetores de entrada (dados)  
Cruzes: vetores codebook de unidades de saída.  
Cor preta: classe 1  
Cor azul: classe 2  
Contorno preto: classificação correta  
Contorno vermelho: classificação errada

# LVQ – Como funciona?





# LVQ – Como funciona?



# Questão de Prova



Implemente o problema apresentado sala de aula em python