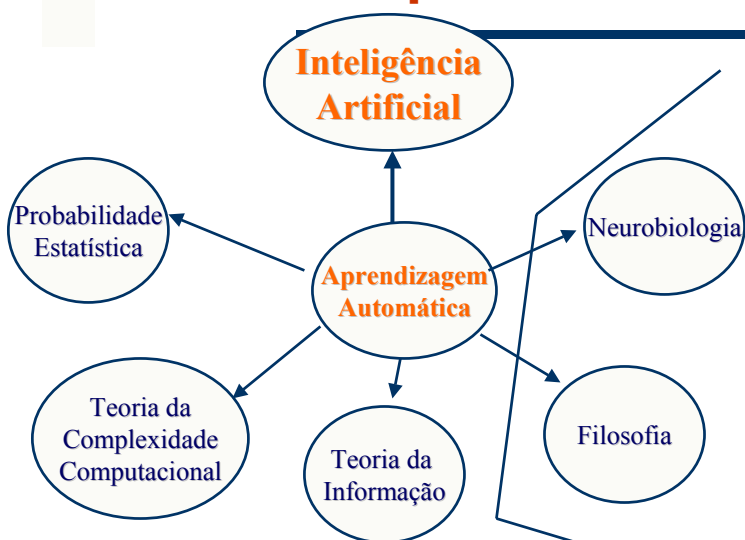

O Essencial sobre Aprendizagem Automática

Seminários de Informática

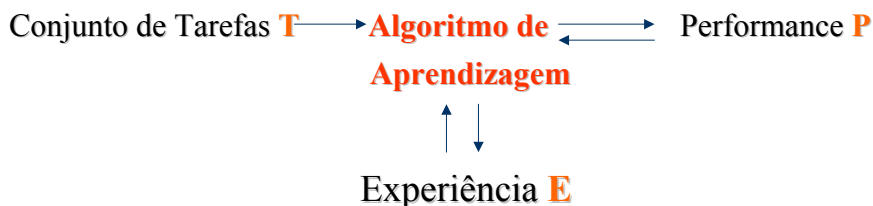
Susana Nascimento
DI- FCT UNL

Área Multidisciplinar



Aprendizagem Automática

o que é?



- **Definição**

[T. Mitchell, 1997]

Um programa **Aprende** a partir da experiência **E**
no que respeita a um conjunto de tarefas **T** e uma medida de desempenho **P**

Se sua *performance* medida por **P** na tarefa **T** num contexto **Z**, aumentar com a experiência **E**.

3

Aprendizagem e Adaptação

- “A learning machine, broadly defined is any device whose actions are influenced by past experiences.” (Nilsson 1965)
- “Any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population.” (Simon 1983)
- “Modification of a behavioral tendency by expertise.” (Webster 1984)
- “An improvement in information processing ability that results from information processing activity.” (Tanimoto 1990)
- “Self-constructing or self-modifying representations of what is being experienced for possible future use” (Michalski, 1990)

4

Exemplo de Aprendizagem

(1)

■ Jogo do Gamão

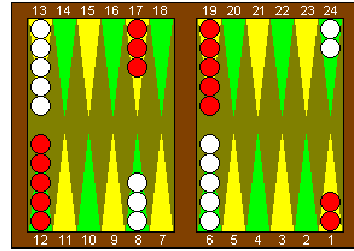
T: jogar do gamão

P: percentagem de jogos ganhos
contra os opositores

E: experiência da prática de jogadas
contra o próprio sistema.

- trained by playing 1.5 million games against itself
- now approximately equal to the best human player

<http://www.research.ibm.com/massive/tcl.html>



5

Exemplo de Aprendizagem

(2)

■ Classificação de texto

T: classificação de textos
em alguma categoria.

P: percentagem de textos
correctamente classificados.

E: base de dados com
alguns textos onde é dada a
classificação correcta dos
mesmos.

Web Site Directory - Sites organized by subject	
Business & Economy B2B, Finance, Shopping, Jobs...	Regional Countries, Regions, US States...
Computers & Internet Internet, WWW, Software, Games...	Society & Culture People, Environment, Religion...
News & Media Newspapers, TV, Radio...	Education College and University, K-12...
Entertainment Movies, Humor, Music...	Arts & Humanities Photography, History, Literature...
Recreation & Sports Sports, Travel, Autos, Outdoors...	Science Animals, Astronomy, Engineering...
Health Diseases, Drugs, Fitness, Medicine...	Social Science Languages, Archaeology, Psychology...
Government Elections, Military, Law, Taxes...	Reference Phone Numbers, Dictionaries, Quotation...

6

Exemplo de Aprendizagem

(3)

- **Condução de um veículo (ALVINN)**

T: conduzir um veículo numa auto-estrada de 4 faixas utilizando sensores de visão.

P: distância média viajada antes de cometer um erro (analisado por um humano).

E: sequência de imagens e instruções de condução registados durante a observação de um condutor humano

7

Aplicações de Impacto

- **Bio-Tecnologia**

- Predicção de estrutura de Proteínas
- *Micro-array gene expression*

- **Predicção de performance de Sistemas de Computadores**

- Sistema de segurança informática

- **Aplicações Bancárias Banking Applications**

- Aplicações de Crédito
- Detecção de Fraudes de cartões de crédito

- **Reconhecimento automático de Caracteres / Voz**

- (I.e. serviço de correio automático)

- **Reconhecimento automático de Voz**

- sistema de assistência automática de telefones

- **Diagnóstico Médico**

- detecção de tumores

- **Aplicações Web**

- classificação de documentos
- Assistência personalizada
- Aprendizagem de preferências do Utilizador

8

Sistema de Aprendizagem

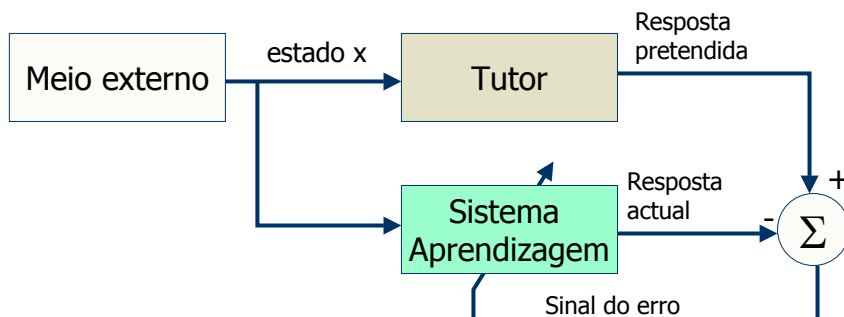
- **Conhecimento a ser aprendido**
 - função objectivo (alvo)
 - Representação da função objectivo
- **Experiência de treino**
- **Processo de Aprendizagem**
 - Algoritmo de aproximação da função
- **Supervisionada *versus* não Supervisionada**

- http://en.wikipedia.org/wiki/Supervised_learning
- http://en.wikipedia.org/wiki/Unsupervised_learning

9

Aprendizagem Supervisionada

- Conhecimento é representado por conjunto de n exemplos
 $\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}$
 - Em que o c_i corresponde à classe do exemplo
- Objectivo: minimizar o erro entre a resposta actual do sistema de aprendizagem, e a resposta 'objectivo'



10

Indução de Árvores de Decisão

- Dado: conjunto de exemplos
- Induzir uma decisão de classificação
 - Indução de regras decisão

▪ http://en.wikipedia.org/wiki/Decision_trees

Um Exemplo

Problema: decidir quando esperar por mesa no restaurante, com base nos seguintes atributos:

1. **Alternate**: existe alguma alternativa próxima? {T, F}
2. **Bar**: existe alguma área de bar confortável para esperar? {T, F}
3. **Fri/Sat**: hoje é Sexta-feira ou Sábado? {T, F}
4. **Hungry**: Estamos com fome? {T, F}
5. **Patrons**: nº de pessoas no restaurante {none, some, full}
6. **Price**: gama de preços {\$, \$\$, \$\$\$}
7. **Raining**: está a chover? {T, F}
8. **Reservation**: foi feita uma reserva? {T, F}
9. **Type**: tipo restaurante {French, Italian, Thai, Burger}
10. **WaitEstimate**: intervalo estimado de espera {0-10, 10-30, 30-60, >60₁₂}

Conjunto de Treino

- Exemplos são descritos por atributos / valores (booleano, discreto, contínuo)
- Atributo classe 'wait' {T, F}

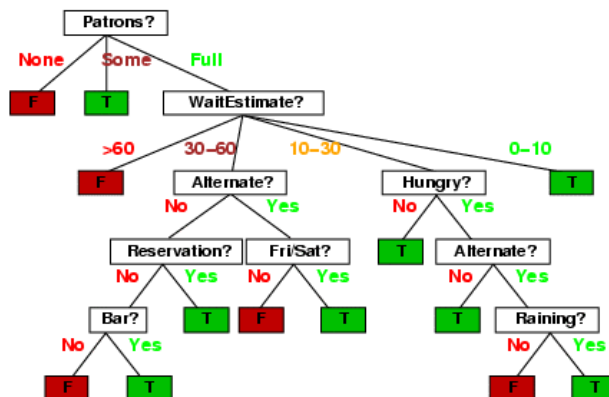
Example	Attributes										Target
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Wait
X_1	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T
X_2	T	F	F	T	Full	\$	F	F	Thai	30-60	F
X_3	F	T	F	F	Some	\$	F	F	Burger	0-10	T
X_4	T	F	T	T	Full	\$	F	F	Thai	10-30	T
X_5	T	F	T	F	Full	\$\$\$	F	T	French	>60	F
X_6	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T
X_7	F	T	F	F	None	\$	T	F	Burger	0-10	F
X_8	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T
X_9	F	T	T	F	Full	\$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F
X_{11}	F	F	F	F	None	\$	F	F	Thai	0-10	F
X_{12}	T	T	T	T	Full	\$	F	F	Burger	30-60	T

- Função de decisão a aprender: 'Wait' = {T, F}
- Classificação de exemplos é positiva (T) ou negativa (F)

13

Árvore: representação de hipóteses

- Uma representação possível das hipóteses de decisão
E.g., árvore final para decidir 'Wait' = {T, F}



14

Aprendizagem de Árvore de Decisão

▪ Objectivo

encontrar uma árvore compacta, que seja consistente com os exemplos do conjunto de dados.

▪ Procedimento

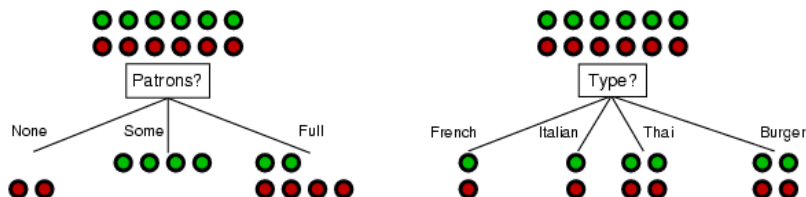
Escolher, recursivamente, o “melhor” atributo para classificar os exemplos, como folha da sub-árvore a expandir

15

Escolha de um “bom” Atributo

▪ Ideia

Um “bom atributo” é aquele que (idealmente) divide o conjunto de exemplos nas classes distintas: “todas positivas” ou “todas negativas”



– O atributo ‘Patrons’ é uma boa escolha

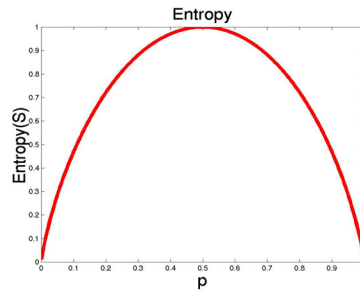
16

Entropia e Ganho de Informação

Seja

- S uma amostra de exemplos de treino
- p_+ a proporção de exemplos positivos
- p_- a proporção de exemplos negativos

$$Entropia(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

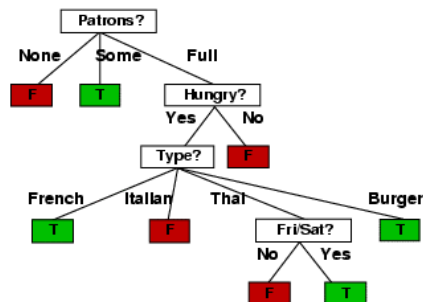


- A entropia é zero quando o conjunto é perfeitamente homogêneo.
- A entropia é um quando o conjunto é totalmente não homogêneo.

17

Conhecimento Aprendido

- Árvore de Decisão aprendida a partir dos 12 exemplos



- Regras de Indução extraídas

Wait(r) \Leftrightarrow Patrons(r , 'some')

- ✓ Patrons(r , 'full') \wedge Hungry(r) \wedge Type(r , 'french')
- ✓ Patrons(r , 'full') \wedge Hungry(r) \wedge Type(r , 'thai') \wedge Fri_Sat(r)
- ✓ Patrons(r , 'full') \wedge Hungry(r) \wedge Type(r , 'burger')

18

Propriedades

- Aproxima funções
 - É capaz de representar qualquer função
 - Booleana, Discreta, contínua
- É eficiente em grandes quantidades de dados
- É um dos métodos mais práticos e mais utilizados em AA
 - Interpretabilidade da Árvore
 - Fácil extrair regras de indução
- Utilizado com sucesso em grande diversidade de problemas
 - Diagnóstico médico, análise de risco de crédito, etc ...
 - Os peritos apreciam !!!

19

Redes Neurais

**Aprendizagem de Pesos para
Aproximação de Funções de forma adaptativa**

- http://en.wikipedia.org/wiki/Artificial_neural_network
- <http://en.wikipedia.org/wiki/Perceptron>

Motivação Biológica

Neurobiologia

Actividade do neurónio é activada/inibida através das ligações a outros neurónios

Números

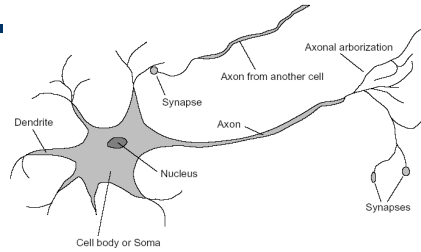
Tempo de comutação: 10^{-3} segundo

Número de neurónios: 10^{11}

Ligações por neurónio: 10^4 (em média)

Tempo de reconhecimento: 10^{-1} segundo

Processo altamente paralelo

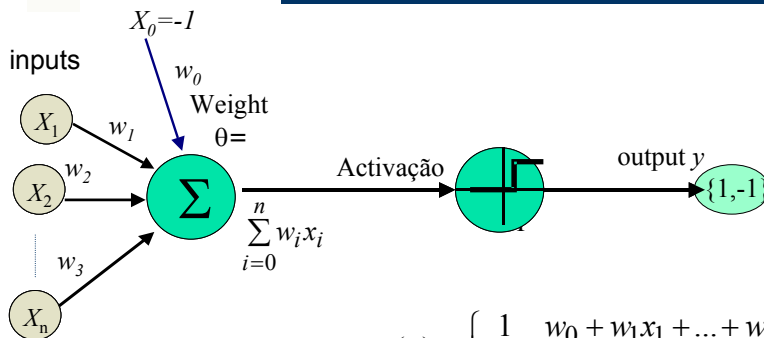


■ Redes Neurais

- Muitas unidades de comutação semelhantes aos neurónios.
- Muitas ligações (pesadas) entre unidades.
- Processo inerentemente paralelo / distribuído
- Ênfase no ajuste automático dos pesos

21

Perceptrão



$$y = g(\mathbf{x}) = \begin{cases} 1 & w_0 + w_1 x_1 + \dots + w_n x_n > 0 \\ -1 & c.c. \end{cases}$$

Combinação linear de atributos + unidade de limiar

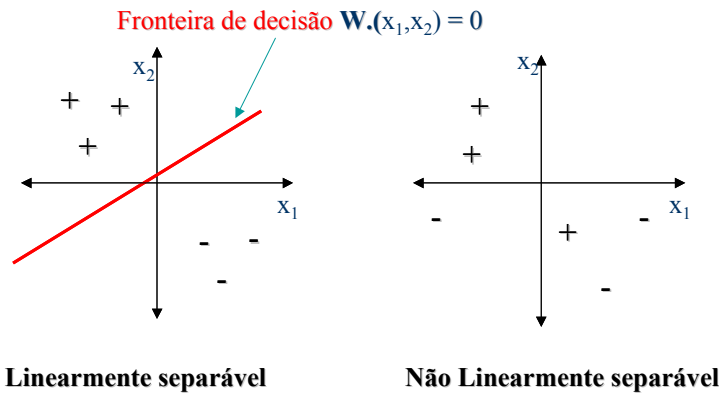
APRENDER um perceptrão significa encontrar os melhores valores para os pesos W .

O espaço de hipóteses de um perceptrão é o espaço de todos os vectores de pesos W .

22

Poder de Representação (1)

Um perceptrão só consegue aprender exemplos que sejam “linearmente separáveis”: os que são perfeitamente separáveis por um hiperplano.

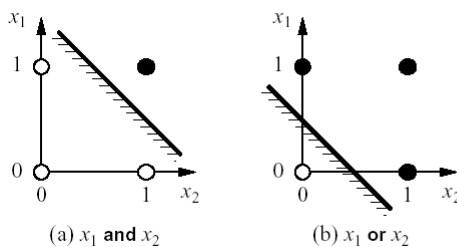


23

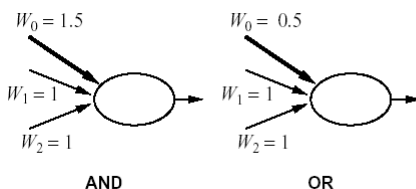
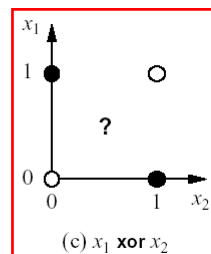
Poder de Representação (2)

- Representa funções booleanas

- fronteira de decisão



XOR Não é linearmente separável



● true
○ false

24

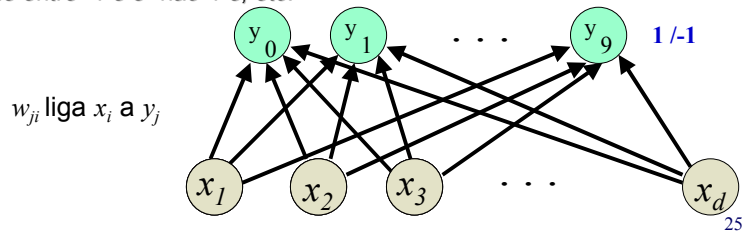
Exemplo com Múltiplas Unidades

- Reconhecimento automático de algarismos manuscritos



- 10 classes : '0', '1', '2', ..., '9'

Unidade correspondente a y_0 distingue entre "0"s e "não-0"s, segunda unidade entre "1"s e "não-1"s, etc.



Problema de Aprendizagem

- Conhecimento a Aprender**

— Vector de pesos $W = (w_1, w_2, \dots, w_n)$ que faz o perceptrão produzir a saída correcta $y = (\pm 1)$ para os exemplos de treino X .

- Algoritmo Iterativo**

- Afectar pesos aleatórios a cada entrada.
- Aplicar iterativamente o perceptrão a cada exemplo, modificando os pesos w sempre que o exemplo seja mal classificado.
- Repetir o processo até que o perceptrão classifique correctamente todos os exemplos.

- Regra de Treino do Perceptrão**

$$w_i \leftarrow w_i + \Delta w_i, \quad \Delta w_i = \eta(t - y)x_i$$

$t = f(x)$ valor da função objectivo

y valor de saída do perceptrão

η ritmo de aprendizagem (i.e. $\eta = 0.1$)

- Regra de Treino Delta**

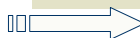
Minimizar o erro quadrático de uma hipótese de treino

$$W = (w_1, w_2, \dots, w_n)$$

$$E(w) = \frac{1}{2} \sum_{d \in D} (t_d - y_d)^2$$

$$w \leftarrow w + \Delta w,$$

$$w_i \leftarrow w_i + \Delta w_i, \quad \Delta w_i = -\eta \sum_{d \in D} (t_d - y_d) x_{id}$$



Treino de Rede Neuronal: 'AND'

- **Dado** conjunto de treino S de exemplos $\{x, t\}$
 - $x = (x_1, x_2, \dots, x_d)$ é o vector de entrada
 - t é a função objectivo
 - Exemplo: \wedge lógico

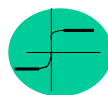
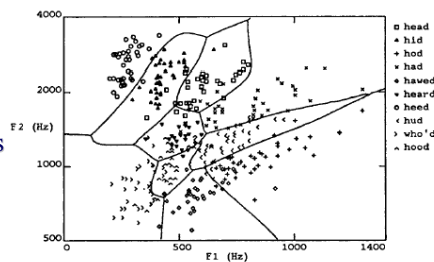
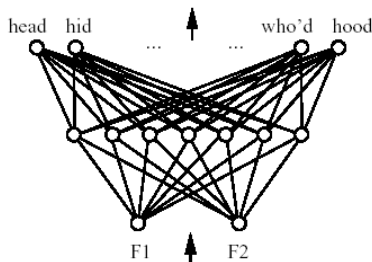
$$S = \{(0,0),0\}, \{(0,1),0\}, \{(1,0),0\}, \{(1,1),1\}$$
- **Processo iterativo**
 1. Apresentar exemplo de treino X ,
 2. comparar saída da rede $y=g(x)$ com o valor original t
 3. ajustar os pesos w_i e limiares θ
- **Regra de Aprendizagem**
 - Aprende conjunto de pesos w e limiares θ da rede, como uma função das entradas x , saída y e função objectivo t .

27

Rede Neuronal Multi-camada

Aprendizagem

- Fronteiras de decisão distintas
- Fronteiras podem ser não lineares



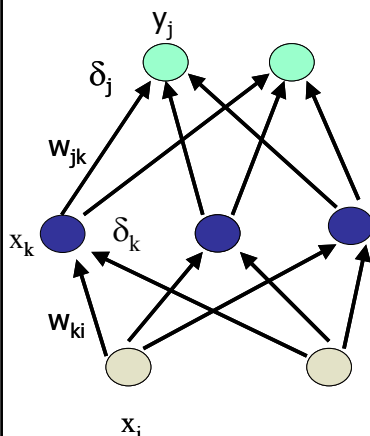
output y

Função logística ou sigmóide

$$y = \sigma(\mathbf{w} \cdot \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w} \cdot \mathbf{x}}}$$

28

Treino: Algoritmo de *Back propagation*



Processo iterativo

Passo Forward Propaga activação da camada de entrada para camada de saída

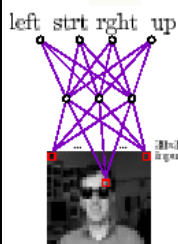
Cada unidade, começando nas de entrada calcula a sua função de activação e transmite-a a todas as unidades a que está ligada, propagando o sinal até às unidades de saída.

Passo Forward Propaga os erros da camada de saída para camada intermédia

Cada unidade de saída compara a sua activação com a saída desejada. O erro é propagado "para trás", para os nós directamente ligados à saída, ajustando os pesos das ligações com base no erro propagado.

29

Reconhecimento de Faces



Learned Weights



Funções a aprender

- direcção para onde olha {E, D, F, C}
- género (M, F)
- se usa óculos de sol (V, F)
- expressão da pessoa (C, T, Z, Neut.)

Rede $960 \times 3 \times 4$

- Treinada pelo backpropagation em 260 imagens
- precisão de classificação de 90% num conj. distinto.



Imagens de entrada típicas

32 imagens por pessoa

Resolução 120×128

Cada pixel é um valor 0–255 (escala de cinzentos)

Neural Networks for Face Recognition

<http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/faces.html>

30

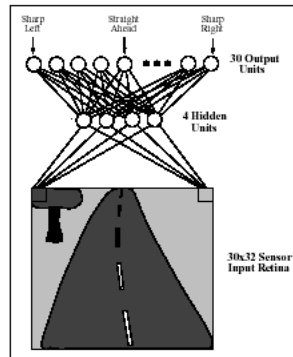
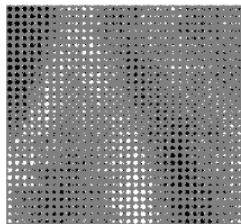
Sistema ALVINN

- **Condução Automática de veículo a 70 mph numa auto-estrada**

Imagem
De camera



Input:
Grelha 30x32
pixeis de
intensidade



Output:

Direcção de viragem do veículo

30x32 weights
into one out of
four hidden
unit

31

Poder expressivo de Rede Neuronal

- **Funções booleanas**

- qualquer função booleana pode ser representada por uma rede com **uma única camada** escondida.
pode exigir um número exponencial de unidades.

- **Funções contínuas**

- Qualquer função contínua limitada **pode ser aproximada**, com um erro arbitrariamente pequeno, por uma rede com uma **camada escondida**.

- **Funções Arbitrárias**

- Qualquer função arbitrária **pode ser aproximada** por uma rede com duas camadas escondidas.

32

Propriedades das Redes Neurais

- ✓ Os exemplos podem ser descritos por grande número de attributes (e.g., pixels de uma image).
- ✓ O valor de saída pode ser discreto, contínuo ou vector de valores discretos e contínuos
- ✓ Dados podem conter erros. Redes neurais são robustas ao ruído.
- A fase de treino da rede é computacionalmente pesada (i.e. 1000-10000 iterações)
 - arquitectura das camadas escondidas não é fácil.
- ✓ A avaliação da rede com novos exemplos é relativamente rápida.
- Uma rede neuronal é tratada como uma “caixa negra”, pelo que a correspondente estrutura não é interpretável.

33

Aprendizagem não Supervisionada

- http://en.wikipedia.org/wiki/Unsupervised_learning
- http://en.wikipedia.org/wiki/Data_clustering
 - http://en.wikipedia.org/wiki/Data_clustering#k-means_clustering

Aprendizagem Não Supervisionada

- Aprendizagem auto-organizada
- Não considera um Tutor
- Objectivo: identificar regularidades nos dados e descobrir grupos de forma automática.
- Aprendizagem competitiva

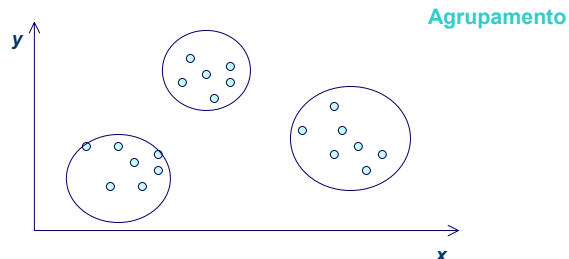


35

O que é Agrupamento?

Dado um conjunto de entidades, organizá-las em grupos (classes)

- forte similaridade dos elementos pertencentes a um mesmo grupo.
- fraca similaridade dos elementos pertencentes a grupos diferentes.



36

Aplicações: exemplos

- Agrupar (*segmentar*) uma base dados de consumidores, com base em padrões de consumo semelhantes.
 - Padrões de consumo diário de electricidade
- Criação de mapas temáticos de SIG, por identificação de grupos espaciais.
- Agrupar perfis de utilizadores de páginas Web, por identificação padrões de preferência de páginas.
- Agrupamento de dados Weblog para identificação de grupos de padrões de acesso.
- **Genomics:** identificar grupos de genes com expressões similares
- ...

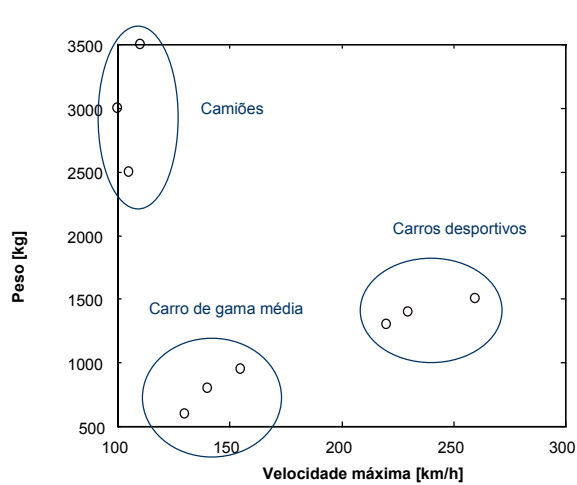
37

Exemplo: Veículos

Veículo	Velocidade máxima (km/h)	Cor	Resistência Ar	Peso (Kg)
V1	220	vermelho	0.30	1300
V2	230	preto	0.32	1400
V3	260	vermelho	0.29	1500
V4	140	cinza	0.35	800
V5	155	azul	0.33	950
V6	130	branco	0.40	600
V7	100	preto	0.50	3000
V8	105	vermelho	0.60	2500
V9	110	cinza	0.55	3500

38

Agrupamento de Veículos

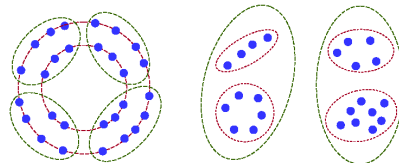


39

Agrupamento vs. Classificação

- Não existe conhecimento *à priori*

- Número de grupos
- Significado dos grupos



- Aprendizagem não Supervisionada

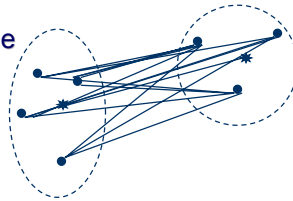
40

Distância entre Grupos

- É necessário definir uma medida de (di)similariedade entre elementos dos grupos.
- **Distância Euclideana**

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2}$$

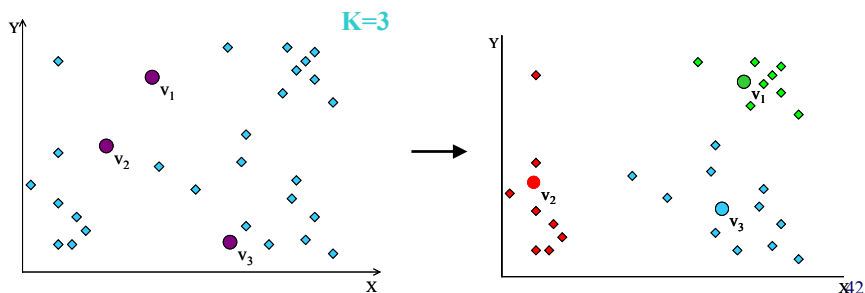
- Existem outras medidas de dissimilariedade
- Distâncias podem ser pesadas



41

Agrupamento por Partição

- Definir o número de grupos, k .
- Definir uma distância entre grupos.
- Os k grupos são construídos iterativamente, passo a passo.
- Resulta num conjunto de grupos das entidades.



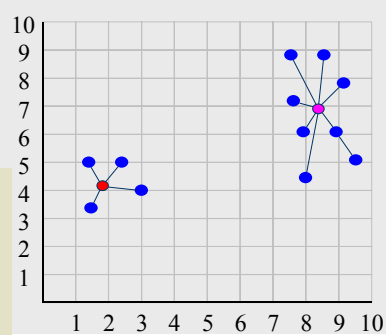
K-Médias (*K-Means*)

- Algoritmo Iterativo que minimiza a distância total das entidades aos centros dos grupos (protótipos).

➤ Iterative distance-based clustering

➤ Steps:

1. Clusters the data into k groups where k is predefined
2. Select k points at random as cluster centers
3. Assign instances to their closest cluster center according to the Euclidean distance function
4. Calculate the centroid or mean of all instances in each cluster (this is the *mean* part)
5. Goto the step 3. Continue until the same points are assigned to each cluster in consecutive rounds.



43

K-Médias

Dado

- Entidades $\{x_1, x_2, \dots, x_n\}$
- Número de grupos K
- Escolher aleatoriamente K protótipos $\{v_1, v_2, \dots, v_k\}$

Repetir

- cada entidade x_i ($i=1, \dots, n$) afecta ao grupo C_k ($k=1, \dots, K$) cujo protótipo v_k está mais próximo.

$$C_k = \{x_1^k, x_2^k, \dots, x_m^k\}$$

$$\text{protótipo } v_k = \frac{1}{m} \sum_{j=1}^m x_j^k$$

- Até** os K protótipos V_k ($k=1, \dots, K$) não se alterarem entre duas iterações consecutivas.

44

Exemplo do Algoritmo K-Médias

Dado:

– {2,4,10,12,3,20,30,11,25}, K=2

■ Escolher aleatoriamente protótipos: $v_1=3, v_2=4$

Iterações

(i) $C_1=\{2,3\}$ $C_2=\{4,10,12,20,30,11,25\}$,
 $v_1=2.5$, $v_2=16$

(ii) $C_1=\{2,3,4\}$ $C_2=\{10,12,20,30,11,25\}$,
 $v_1=3$, $v_2=18$

(iii) $C_1=\{2,3,4,10\}$, $C_2=\{12,20,30,11,25\}$,
 $v_1=4.75$, $v_2=19.6$

(iv) $C_1=\{2,3,4,10,11,12\}$, $C_2=\{20,30,25\}$,
 $v_1=7$, $v_2=25$

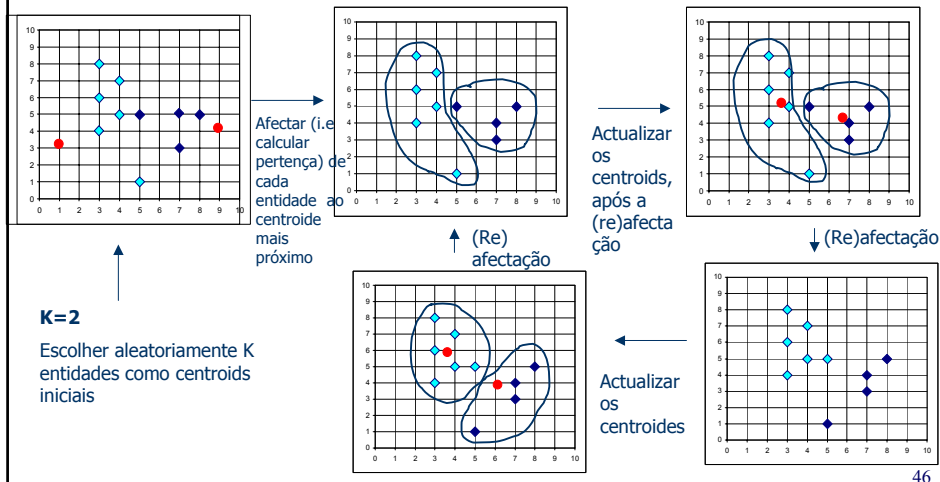
v. $C_1=\{2,3,4,10,11,12\}$, $C_2=\{20,30,25\}$,
 $v_1=7$, $v_2=25$

iterações (iv) e (v) iguais !

45

Passos do Algoritmo K-Médias

Exemplo



Observações

▪ Agrupamento por Partição

- Prático e computacionalmente eficiente (algoritmos do tipo K -médias convergem rapidamente).
- Tendência para definir grupos com forma hiper-esférica (i.e. distância Euclideana).
- Os resultados podem variar significativamente com base na escolha dos protótipos iniciais.
 - Iterar muitas vezes a partir de protótipos iniciais distintos.
- Os protótipos são elementos representativos do conceito associado ao grupo.
- A determinação do número de grupos é fundamental.

47

Para saber mais ...

▪ Wikipedia

- http://en.wikipedia.org/wiki/Main_Page

▪ <http://www.aaai.org/AITopics/html/machine.html>

Does Machine Learning Really Work? By Tom M. Mitchell. AI Magazine, 18(3): Fall 1997.
(<http://www.aaai.org/Library/Magazine/Vol18/18-03/vol18-03.html>)

Introduction to Machine Learning - Draft of Incomplete Notes. By Nils J. Nilsson.
(<http://robotics.stanford.edu/people/nilsson/mlbook.html>)

Machine learns games 'like a human.' By Will Knight. New Scientist News (January 24, 2005).
(<http://www.newscientist.com/article.ns?id=dn6914>)

48