IC 22/23

Exercícios Resolvidos de Exames Anteriores

Considere uma rede neuronal cuja saída é determinada pela expressão:

$$y = \sum_{i=1}^{N} w_i x_i$$

e em que os pesos são calculados de acordo com a aprendizagem de Hebb supervisionada.

a) Como é designada esta rede neuronal? Especifique a fórmula de atualização dos pesos.
b) Será possível atingir um erro de treino nulo? Justifique.
c) Qual a vantagem em adicionar mais uma camada a esta rede?
d) Explique a relação do método "Pseudoinversa" com a aprendizagem de Hebb, e em que circunstâncias pode ser utilizado.

a. Rede "Adaline" (arquitetura da rede representada no slide 2 do cap.5).
Atualização de pesos pela regra de Hebb supervisionada é dada por:

$$w_{ij}^{new} = w_{ij}^{old} + t_{iq} p_{jq}$$

(slide 3)

b. Sim, caso os padrões de entrada sejas ortogonais (perpendiculares) entre si. (slide 4 cap. 5).

c. Não. Sendo uma rede linear de uma camada é "equivalente" a uma rede linear com "N" camadas. Uma rede linear com várias camadas continua a "ser" uma combinação linear dos parâmetros de entrada (ver Neural Network Design - Cap. 7 e demonstraçºao apresentada no quadro).

d. Consultar slide 8:

Regra de Hebb

$$\mathbf{W} = \mathbf{T}\mathbf{P}^T$$

Pseudoinversa

$$\mathbf{W} = \mathbf{T}\mathbf{P}^+$$

$$\mathbf{P}^+ = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T$$

Útil quando não existe inversa da matriz de" entradas" **P**. Sendo assim, a função de otimização do erro, F(W) pode ser minimizada usando o método da pseudo-inversa.

Pretende-se um sistema de inferência difusa que controle a velocidade de um veículo autónomo, cujo limite de velocidade é de 50 km/h, com base nas variáveis "erro", diferença entre a velocidade desejada e a velocidade atual do veículo; e a "variação do erro". A variável de controlo consiste na variação da força a aplicar. Considere a variável força definida na gama [0 100%].

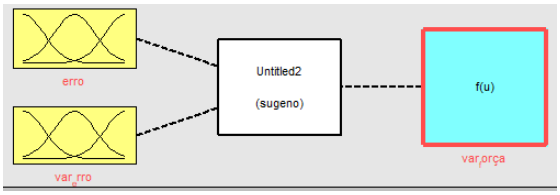
a) Projete um sistema difuso de Sugeno, considerando três termos linguísticos por variável.
b) Calcule a força a aplicar considerando uma velocidade de cruzeiro desejada de 2 km/h e uma velocidade atual de 7 km/h. Use o método de inferência MAX-MIN e o método das alturas para desfuzificação.

a. Variáveis de entrada:

- "erro" (erro=velocidade desejada-velocidade atual)
- "variação do erro" (*var. erro(k)= erro(k)-erro(k-1)*)

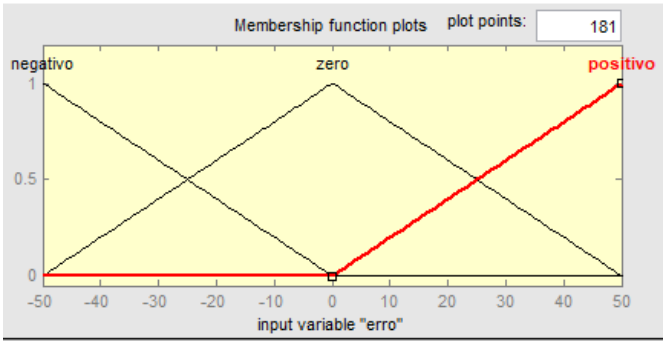
Saída:

- variação da força a aplicar

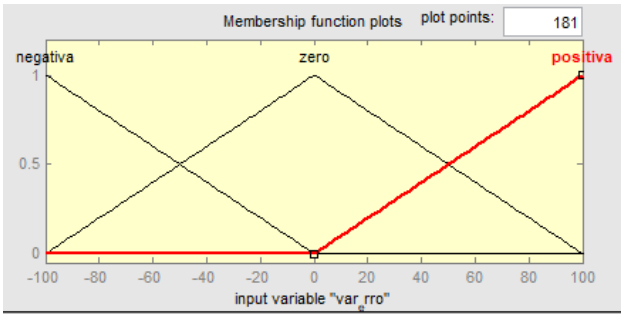


Considerando três termos linguísticos por variável:

Erro – varia entre -50 e +50 (limite de velocidade é de 50 km/h)



Variação do Erro – considera-se variação entre -100 e -100



Habitualmente também podemos considerar uma gama de variação menor, por exemplo igual ao erro, que neste caso seria [-50 +50]. Isto deve-se ao facto de que sendo um sistema dinâmico não existem mudanças bruscas na velocidade.

Gama de Variação da Força (-100 +100)

Sendo um sistema de Sugeno, os consequentes são constantes ou funções lineares. Considerando neste caso consequentes constantes, podem ser definidos por:

Variação Positiva = +50
Variação Nula = 0
Variação Negativa = -50

Base de regras:

Variação da Força:

	Erro = N	**Erro =Z**	**Erro=P**
VE=N	-50	-50	0
VE=Z	-50	0	+50
VE=P	-0	+50	+50

b. Erro= Referência –Velocidade = 2-7 = -5

Var_Erro= Erro - Erro_Anterior=-5-0=-5 (considerando que anteriormente não haveria erro!)

Fuzificação (P,Z,N):
Erro_Fuzificado = (0;0.9;0.1)
VErro_Fuzificado = (0;0.95;0.05)

Assim, “são ativadas” as seguintes quatro regras (na resolução completa deve-se apresentar todos os cálculos para os valores dos antecedentes pelo método MAX-MIN):

	Erro = N	Erro =Z
VE=N	-50	-50
VE=Z	-50	0

Pela regra de Sugeno:

VarForça= (0,1*(-50) + 0,9*(0) + 0,05*(-50) + 0,05*(-50)) / (0,1+0,9+0,05+0,05) = -9,09

Pretende-se aplicar o PSO, numa topologia em roda, para ajustar o peso associado a três regras do sistema difuso. Num possível cenário de utilização, pretende-se que a partícula "1" comunique com as restantes.

a) Considere as coordenadas das posições atuais: x_1=(2,1,3); x_2=(1,2,3); x_3=(3,3,3). As suas melhores posições individuais são respetivamente (2,1,2), (1,2,1), (2,2,3). A melhor posição global é (1,2,1). As velocidades atuais são respetivamente de (1,0,0), (0,1,1) e (1,0,0). Determine as próximas posições das três partículas, assumindo constantes cognitiva, social de um, inércia de 0,5 e velocidade máxima de 5.
b) Apresente duas diferenças entre o algoritmo PSO e um algoritmo de melhoramento iterativo trepa-colinas

a. Posições atuais:
x1=(2,1,3); x2=(1,2,3); x3=(3,3,3)

x1best = (2,1,2),
x2best = (1,2,1),
x3best = (2,2,3)

Gbest = (1,2,1)

V1(k) = (1,0,0)
V2(k) = (0,1,1)
V3(k) = (1,0,0)

constantes cognitiva = 1
constante social =1
inércia = 0,5
velocidade máxima = 5

De acordo com a fórmula apresentada nos slides 5-7 (PSO) e numa topologia em roda, onde apenas a "1" comunica com as restantes:

(falta apresentar todos os passos para o cálculo dos valores…)

v1(k+1) = [-0,5 1 -3]
x1(k+1) = [1,5 2 0]

v2(k+1) = [0 0,5 -3,5]
x2(k+1) = [1 2,5 -0,5]

Para partícula 3, como não "comunica" com a "2", considerando gbest=lbest de 1:

V3(k+1) = [-1,5 -3 -1]
X3(k+1) = [1,5 0 2]

b. Duas diferenças entre o algoritmo PSO e o trepa-colinas:
   - O PSO é pesquisa global (trabalha com várias soluções em simultâneo) enquanto trepa-colinas é local (trabalha com apenas uma solução em cada iteração).
   - O trepa-colinas para num ótimo local (quando todos os vizinhos são piores) enquanto o PSO, na versão "Global" atualiza as posições com base no melhor de entre todas as atuais, o que permite ultrapassar "ótimos locais".

Considere que é necessário determinar a rota de menor distância percorrida por um veículo, devendo passar obrigatoriamente por cinco locais, com distâncias representadas na tabela a) da Figura 1. Pretende-se resolver o problema pelo algoritmo ACO.

a) Proponha a representação da solução e função de avaliação.
b) Determine a probabilidade de escolha do próximo local a visitar para uma formiga partindo do nó "C". A tabela a) indica as distâncias entre os locais e os valores iniciais de feromona estão representados na tabela b) da Figura 1. Considere os parâmetros de influência da taxa de feromona e heurísticas iguais a 1.
c) De que forma poderia aplicar o método da roleta para a seleção do próximo nó a visitar.
d) Suponha que após a primeira iteração, a rota determinada por uma formiga foi "CBAED". Admitindo que a aresta B-D está apenas incluída nesta rota, determine a nova concentração de feromona nesta aresta, usando uma taxa de evaporação de 0.8.

e) Para problemas de *clustering*, o algoritmo ACO tende a criar mais agrupamentos do que os habitualmente necessários. Qual a razão para este comportamento? Descreva uma estratégia para solucionar este problema.

0	2	5	5	3
2	0	2	4	6
5	2	0	5	5
5	4	5	0	6
3	6	5	6	0

(a) Distâncias

0	0,1	0,1	0,2	0,2
0,1	0	0,1	0,2	0,3
0,1	0,1	0	0,5	0,1
0,2	0,2	0,5	0	0,6
0,2	0,3	0,1	0,6	0

(b) taxas de feromona

Figura 1. Valores de distâncias entre nós e taxas atuais de feromona (b).

a. Representação da solução = **array de 5 inteiros**.
Por exemplo a rota 1-2-3-4-5 é representada pelo array S=[1 2 3 4 5].

função de avaliação = **Distância total percorrida naquela rota**
= dist(S(1),S(2))+ dist(S(2),S(3))+ dist(S(3),S(4))+ dist(S(4),S(5))+ dist(S(5),S(1))

b. Probabilidade de escolha do próximo local a visitar para uma formiga partindo do nó "C":

De acordo com fórmulas do slide 9 (ACO):

$$\emptyset_{ij,k}(t) = \begin{cases} \dfrac{\tau_{ij}(t)^{\alpha}\,\eta_{ij}^{\beta}}{\sum_{c \in C_{i,k}} \tau_{ic}(t)^{\alpha}\eta_{ic}^{\beta}}, & se\ j \in C_{i,k} \\ 0, & caso\ contrário \end{cases}$$

P(3,1)= 10.5%

P(3,2)= 26.3%

P(3,4)= 52.6%

P(3,5)=10,5%

(na resolução completa deve apresentar-se todos os passos necessários para o cálculo destas probabilidades)

c. Pelo método da roleta:

As áreas correspondentes na roleta são:

R1=10,5
R2=36,8
R3=89,4
R4=100

Geramos um valor aleatório entre 0 e 1. Por exemplo:

```
>> rand

ans =

    0.8147

>>
```

Neste caso escolheríamos a "cidade 4" (pois corresponde ao intervalo entre 36,8 e 89,4).

d. Suponha que após a primeira iteração, a rota determinada por uma formiga foi "**CBAED**". Admitindo que a aresta **B-A** está apenas incluída nesta rota, determine a nova concentração de feromona nesta aresta, usando uma taxa de evaporação de 0.8

Usando as fórmulas do slide 10 (ACO):

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t)$$

$$\Delta\tau_{ij,k}(t)\begin{cases} \frac{Q}{L_k(t)} & se\ (i,j) \in T_k(t) \\ 0 & caso\ contrário \end{cases}$$

Sendo Q o valor ótimo, e sendo desconhecido, podemos admitir o valor Q=5 (nenhuma ligação tem valor inferior a 1).
L= Comprimento da rota **CBAED =** 2+2+3+6+5=18
Novo taxa de feromona na ligação "B-A" = (1-0,8)*0,1+(5/18)=0,29

e. Qual a razão para este comportamento?

As formigas ao não possuírem memória para guardar a posição dos "grupos" anteriormente formados, corre-se o risco de criar um ou mais grupos para a mesma classe.

f. Descreva uma estratégia para solucionar este problema.

Podemos definir velocidades diferentes para as formigas ou usar memória de curto prazo.

Considere que é necessário determinar a rota de menor distância percorrida pelo drone, devendo passar obrigatoriamente por quatro pontos de entrega, como representado no grafo da Figura 2. Pretende-se resolver o problema pelo algoritmo ACO.

a) Determine a probabilidade de escolha da próximo local a visitar, para uma formiga partindo do nó "B". O grafo indica as distâncias entre os locais e os valores iniciais de feromona. Considere os parâmetros de influência da taxa de feromona e heurísticas iguais a 1.

b) Aplique o método da roleta para escolher a cidade a visitar, considerando um valor gerado aleatoriamente de 0,5.

c) Considerando que após uma iteração foram determinadas as seguintes rotas: [ACBD]; [BADC]; [CABD] e [DABC]. Determine a nova taxa de feromona na aresta "BC", considerando uma taxa de evaporação de 0,1.

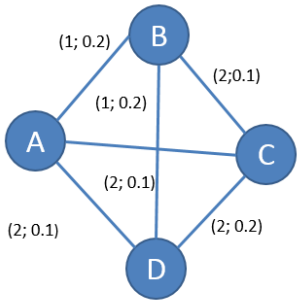


Figura 2. Grafo do problema e taxas atuais de feromona.

a)

P(B,A)=(0.2*1)/((0.2*1)+ (0.2*1)+ (0.1*0.5))=0.2/(0.2+0.2+0.05)=0.2/0.45=44,44%
P(B,C)=(0.1*0.5)/0.45=0.05/0.45=11,11%
P(B,D)=(0.2*1)/0.45=0.2*0.45=44,44%

(0,666667
0,166667
0,166667)

b)

R(A)=1/3=0.33

R(C)=2/3=0.55
R(D)=100=1
Com rand=0.5, 0.33<0.5<0.55, escolhe C.

c) Custo [ACBD]=2+2+1+2=7
Custo [BADC]=; 1+2+2+2=7
Custo [DABC]=2+1+2+2=7

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t)$$

Nova Taxa=(0.9*0.1)+(1/7+1/7+1/7)=0.09+3/7=0.09+0.43=0.52

Pretende-se uma rede neuronal para classificação de dígitos (0 .. 9). Considere um dígito representado numa matriz binária de dimensão 4x4, como representado na Figura 1(a).

a) Proponha uma possível arquitetura de uma rede neuronal MLP para resolver o problema de classificação. Especifique número de entradas, camadas, número de saídas e seu significado.
b) Para uma rede CNN, considerando o filtro representado na Figura 1(b), determine o "feature map" resultante, com "stride" de 1 e sem "padding".
c) Com a aplicação do filtro representado, seria possível obter um "feature map" com a mesma dimensão da imagem de entrada? Justifique.

0	0	1	0
0	1	1	0
0	0	1	0
0	0	1	0

Figura 1(a)- Imagem a classificar

1	0	0
0	1	0
0	0	1

Figura 1(b)- Filtro

a) 16 Entradas (representa a matriz com representação do dígito, 4*4 valores binários) e 10 saídas (dígitos de 0.9). O número de camadas internas e número de neurónios deve ser ajustado com base nos resultados de treino.

b) Fm(1,1)=0*1+1*1+1*1+1*1=2
Fm(1,2)=0*1+1*1+1*1+0*1=1
Fm(2,1)=0*1+0*1+1*1+1*1=1
Fm(2,2)=1*1+1*1+1*1+0*1=2

Corresponde à soma dos valores na diagonal principal da sub-imagem.

2	1
1	2

c) Sim, considerando padding como representado, teríamos um "feature map" de 4*4 (16 filtros).

Imagem com padding:

0	0	0	0	0	0
0	0	0	1	0	0
0	0	1	1	0	0
0	0	0	1	0	0
0	0	0	1	0	0
0	0	0	0	0	0

Feature map resultante é de 4*4:

1	1	1	0
0	2	1	1
0	1	2	1
0	0	1	1

Pretende-se aplicar o algoritmo PSO otimizar dois parâmetros de uma rede neuronal. Considere as coordenadas das posições atuais: x1=(1,1); x2=(0,1); x3=(1,2). As suas melhores posições individuais são respetivamente (1,0,), (2,1), (1,2). A melhor posição global é (1,2). As velocidades atuais são respetivamente de (1,1), (0,1) e (1,0). Assumindo constantes cognitiva e social de valor igual a 1, inércia de 0,5 e velocidade máxima de 10.

a) Determine a próxima posição da partícula X1, para a versão individual best do PSO.
b) Determine a próxima posições da partícula X2, considerando agora uma topologia em estrela.
c) Como procederia para obter os melhores parâmetros da rede neuronal numa pesquisa em grelha? Compare com o PSO.

a) v1=0.5[1 1]+1*([1 0]-[1 1])=[0.5 0.5]+[0 -1]=[0.5 -0.5]
x1=[1 1]+[0.5 -0.5]=[1.5 0.5]

b) v2=0.5[0 1]+1*([1 2]-[0 1])+ 1*([2 1]-[0 1])=[0 0.5]+[1 1]+ [2 0]=[3 1.5]
x2=[0 1]+[3 1.5]=[3 2.5]

c)

Define-se a gama para a dim1, por exemplo X(1)=(0:0.1:10), e para a dim2, X(2)=(0:0.1:10). Uma pesquisa em grelha iria avaliar todas as combinações possíveis, que neste caso seriam 100*100=10000!

O PSO permite otimizar o tempo de pesquisa, não necessita de avaliar todas as posições. Especialmente útil para espaços de dimensão elevada onde não é possível avaliar todas as possibilidades em tempo útil (no caso anterior teríamos de treinar 10000 redes neuronais!)

Indique se as seguintes afirmações são verdadeiras ou falsas e justifique:

a) Uma rede neuro-difusa necessita de incorporar conhecimento pericial.

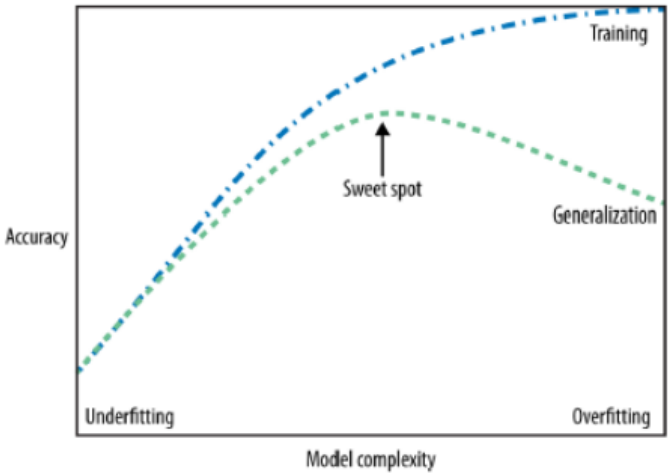
Falso.

Uma rede neuro-difusa (por exemplo uma rede ANFIS), sendo também uma rede neuronal, poder ser treinada apenas com base em dados (Aprendizagem Automática), sem necessidade de inclusão de conhecimento por parte de um humano - perito no domínio da aplicação.

b) A rede com melhor capacidade de generalização é aquela que possui um maior número de neurónios e camadas (parâmetros).

Falso.

Quando o número de parâmetros da rede neuronal é demasiado elevado, existe um sobre-ajustamento aos dados de treino (overfitting) e assim a rede perde capacidade de generalização (erro de teste aumenta!).



c) Na variante *local best* do algoritmo PSO, o comportamento das partículas é determinado apenas pela sua própria experiência.

Falso.
Na versão "local best", as partículas são influenciadas pela sua própria experiência e pelas partículas dentro do seu raio de vizinhança.

d) A aplicação da constante de inércia ao algoritmo PSO garante sempre a convergência do algoritmo.

Falso.
Diversos estudos comprovam que o PSO não converge para qualquer combinação de parâmetros, devendo-se respeitar a relação da fórmula:

$$\frac{1}{2}(\rho_1 + \rho_2) - 1 < \varphi \leq 1$$

e) Na aplicação do algoritmo ACO ao problema do TSP, se usarmos apenas informação sobre a taxa de feromona (sem considerar heurísticas), aumenta a probabilidade de existir convergência para rotas sub-óptimas.

Verdadeiro.
Se "β=0", será apenas usada a informação sobre as feromonas, e pode existir convergência para rotas sub-óptimas (ótimos locais)

f) O número de parâmetros de uma rede CNN é bastante superior a uma rede MLP para classificação de imagens (com a mesma dimensão).

Falso.
A operação de convolução (aplicação dos filtros) permite reduzir consideravelmente o número de parâmetros.
Exemplo:

para imagens de dimensão "150 × 100":

Se usar uma rede MLP com o mesmo número de neurónios da entrada:
Número de parâmetros = 15000*15000= 225 000 000!

Se usar uma rede CNN:
Filtros 5*5 produzindo "200 features maps" de 150*100 stride=1
Número de parâmetros = (5*5 + 1) × 200 (com bias) = 5200 parâmetros!!