

INATEL – Instituto Nacional de Telecomunicações

C210 – Inteligência Computacional

Profa. Victoria Dala Pegorara Souto

Aula 7 – Redes Neurais

1. Defina o que é uma Rede Neural Artificial.
2. Quais as principais semelhanças e diferenças entre um cérebro humano e um computador convencional? Dê sua resposta em termos das capacidades de processamento e aprendizado.
3. Qual fato apresentado em 1969 por Minsky & Papert implicou no congelamento da área de pesquisa de redes neurais? Como este problema foi resolvido posteriormente, por Rumelhart *et al* em 1986?
4. Cite ao menos três áreas de aplicações das redes neurais artificiais.
5. Cite as arquiteturas básicas de redes neurais, apontando as principais características de cada.
6. Quais são os elementos fundamentais que compõem um neurônio artificial? Ilustre, nomeie e explique a função de cada componente.
7. Associe os termos da esquerda às definições da direita:

| | | | |
|---|-------------------------------------|-----|-----------------------|
| A | $\{x_1, x_2, \dots, x_m\}$ | () | Pesos sinápticos |
| B | $\{w_{k1}, w_{k2}, \dots, w_{km}\}$ | () | Potencial de ativação |
| C | $\{\Sigma\}$ | () | Combinador linear |
| D | $\{\theta\}$ ou $\{b_k\}$ | () | Sinais de entrada |
| E | $\{v_k\}$ | () | Sinal de saída |
| F | $\{\varphi\}$ | () | Função de ativação |
| G | $\{y_k\}$ | () | Limiar de ativação |

8. Quais fatores influenciam para que um potencial de ativação seja inibitório ou excitatório?
9. Explique brevemente as principais funções de ativação, levantando semelhanças e diferenças.

10. Quando é possível afirmar que uma rede neural “aprendeu” um determinado problema?
11. Como é chamada cada etapa de apresentação completa dos dados de treinamento?
12. Como é possível medir a diferença entre um par de dados de entrada? E a similaridade?
13. Explique as principais diferenças entre treinamento supervisionado e não supervisionado.

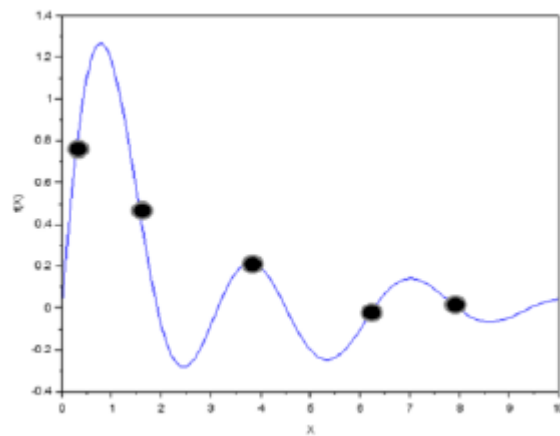
Redes Perceptron

1. Explique por que o Perceptron é um classificador linear.
2. Após diversos processos independentes de treinamento, pode-se afirmar que os vetores de pesos sinápticos ajustados de cada rede Perceptron são iguais (assumindo uma superfície de erro com um único ponto mínimo)? Justifique.
3. É possível utilizar uma rede Perceptron para descobrir se os padrões de um problema de classificação são linearmente separáveis? Se sim, explique qual seria a estratégia.
4. Imagine que dois pesquisadores independentes estão utilizando redes Perceptron para mapear o mesmo problema de classificação de padrões. Responda às questões:
 - a. É correto afirmar que ambas as redes convergirão com o mesmo número de épocas?
 - b. Considere que ambas as redes já estão devidamente treinadas. Para um conjunto contendo 10 novas amostras que devem ser identificadas, explique se os resultados produzidos por ambas as redes serão os mesmos.
5. Considere um problema de classificação de padrões linearmente separável composto de 50 amostras. Em determinada época e_n de treinamento observou-se que somente para uma dessas amostras a rede não estava produzindo a resposta desejada. Neste caso, é necessário apresentar novamente todas as 50 amostras na próxima época (e_{n+1}) de treinamento? Justifique.
6. Um neurônio j recebe os seguintes valores: $\theta = 10$, $w_\theta = 0,8$ e $x = \begin{bmatrix} -20 & 4 & 2 \\ 7 & 15 & -3 \end{bmatrix}$. Seus pesos sinápticos iniciais são: $w = [9,2 - 1,0 - 0,9]$. Faça o que se pede:
 - a. Utilizando a Lei de Hebb, com $\eta = 0,5$, $d = [+1 - 1]^T$ e função de ativação Heaviside simétrica, efetue 2 épocas de treinamento e escreva o vetor de pesos sinápticos ajustados (final).
 - b. Com os pesos ajustados, calcule o resultado do combinador linear (v_k) do neurônio para uma nova entrada $x^{(3)} = [-12 - 68]$.
 - c. Calcule a saída $y^{(3)}$ com função de ativação linear $\varphi(v) = v$.
 - d. Calcule a saída $y^{(3)}$ com função de ativação heaviside $\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases}$.

7. Sobre a taxa de aprendizagem (η), responda às questões:
- Qual a faixa de valores normalmente utilizada para η ?
 - Um valor de η muito pequeno pode afetar o processo de convergência? Explique.
 - Um valor de η muito grande pode afetar o processo de convergência? Explique.

Aula 7 – Redes Adaline

- O que significa o nome “ADALINE”?
- Explique a(s) diferença(s) entre as redes Perceptron e ADALINE, no que diz respeito à:
 - Etapa de treinamento
 - Etapa de operação
- O que o Erro Quadrático Médio (EQM) representa? Qual sua tendência ao longo das épocas de treinamento? Explique.
- Explique se as instabilidades que eventualmente se manifestam nas redes Perceptron devido à adoção de valores inconsistentes para a taxa de aprendizado (η) também podem ocorrer quando utilizada a regra Delta no processo de treinamento.
- Em relação ao critério de parada para o processo de convergência da rede ADALINE, discorra se há realmente a necessidade de aplicação do operador módulo sobre a diferença do erro quadrático médio entre duas épocas de treinamento sucessivas.
- Uma rede ADALINE com quatro entradas possui apresenta os seguintes parâmetros: $\theta = -1$, $w_\theta = -0,3$ e $w = [0,8 \ -0,5 \ -0,1 \ 1]$.
 - Considerando $\eta = 0,5$, $\varepsilon = 0,1$, $x = \begin{bmatrix} 0,2 & -0,5 & 0,8 & 1,2 \\ -0,3 & 0,9 & 1,1 & -0,7 \end{bmatrix}$ e $d = [+1 \ -1]$, efetue 2 épocas de treinamento com a Regra Delta e escreva os pesos sinápticos ajustados.
 - Utilizando a função de transferência Heaviside simétrica, calcule a saída da rede (y) para:
 - $x = [0 \ 0 \ 0 \ 0]$
 - $x = [0 \ 1 \ 0 \ 1]$
 - $x = [1 \ 0 \ 1 \ 0]$
 - $x = [1 \ 1 \ 1 \ 1]$
- Considerando a função do erro (eixo y) em relação aos pesos (eixo x) mostrada no gráfico, indique o sinal do valor do gradiente para cada um dos pontos indicados.



8. Escreva sobre um perigo que pode ocorrer no treinamento de uma rede ADALINE sobre um problema que contenha vários mínimos locais.
9. Após diversos processos independentes de treinamento, pode-se afirmar que os vetores de pesos sinápticos ajustados de cada rede ADALINE são iguais (assumindo uma superfície de erro com um único ponto mínimo)? Justifique.

Redes Perceptron de Múltiplas Camadas

1. Quais as principais diferenças entre as redes SLP (Single Layer) e as MLP (Multilayer Perceptron)?
2. Cite problemas onde as redes MLP são aplicáveis. Considere tanto aspectos técnicos quanto práticos.
3. Desenhe a arquitetura de uma rede MLP, identificando e nomeando seus principais componentes.
4. Explique, com suas palavras, o funcionamento das fases *forward* (para frente) e *backward* (para trás) do algoritmo *backpropagation*, destacando também a finalidade de cada um.
5. Que características as funções de ativação devem possuir para que possam ser utilizadas em uma rede Perceptron multicamadas? Explique e cite as funções mais comumente utilizadas.
6. A ordem de apresentação dos dados durante o aprendizado influencia no desempenho da rede? Se sim, como isto pode ser resolvido? Explique.
7. No treinamento de uma MLP, qual a diferença entre a atualização instantânea e a por lotes?
8. Em que situações pode ocorrer *underfitting*? No que isso implica? Como isto pode ser resolvido?
9. Em que situações pode ocorrer *overfitting*? No que isso implica? Como isto pode ser resolvido?

10. Explique a técnica de validação cruzada, relevante no contexto de treinamento de redes neurais.