**INATEL – Instituto Nacional de Telecomunicações**

C210 – Inteligência Computacional

Prof. Me. Marcelo Vinícius Cysneiros Aragão

AULA 10 - Redes Perceptron de Múltiplas Camadas

**1. Quais as principais diferenças entre as redes SLP (Single Layer) e as MLP (Multilayer Perceptron)?**

**R-** A presença de uma ou mais camadas neurais escondidas e o uso de funções de ativação não-lineares.

**2. Cite problemas onde as redes MLP são aplicáveis. Considere tanto aspectos técnicos quanto práticos.**

**R-** Podem ser aplicadas em diversos tipos de problemas, por exemplo:

Aproximação universal de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais, otimização de sistemas.

**3. Desenhe a arquitetura de uma rede MLP, identificando e nomeando seus principais componentes.**

**R-** O

**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.**

**R-** O treinamento com o algoritmo backpropagation é feito em duas fases: forward (cálculo do erro) e backward (correção dos pesos sinápticos).

1. O passo para frente é onde as entradas são passadas através da rede e as previsões de saída obtidas (essa etapa também é conhecida como fase de propagação).

2. O passo para trás é onde calculamos o gradiente da função de perda na camada final da rede e usamos esse gradiente para aplicar recursivamente a regra da cadeia (chain rule) para atualizar os pesos em nossa rede (etapa também conhecida como fase de atualização de pesos ou retro propagação).

**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.**

**R-** As funções de ativação devem ser contínuas e diferenciável em todo o seu domínio.

**6. Como o termo de momentum pode contribuir no treinamento de uma rede neural?**

**R-** O momentum (0 < α < 0.9) permite a travessia de platôs com mais facilidade devido à “inércia”.

Quando a solução atual estiver longe da final, a variação na direção oposta ao gradiente do EQM de iterações sucessivas será grande

→ grande contribuição do momentum

Quando a solução atual estiver próxima da solução final, as variações nas matrizes de pesos serão então bem ínfimas, pois o EQM entre duas iterações sucessivas será baixa

→ baixa contribuição do momentum

**7. A ordem de apresentação dos dados durante o aprendizado influencia no desempenho da rede? Se sim, como isto pode resolvido? Explique.**

**R-** Sim, a ordem de apresentação dos dados durante o aprendizado pode influenciar o desempenho da rede. Uma forma de evitar isto é apresentá-los de forma aleatória.

* O conjunto de dados deve ser dividido em duas partes:

a) aprendizado

b) testes (validação e testes finais)

* Se o número total de dados é pequeno, deve-se privilegiar o conjunto de aprendizado.

**8. No treinamento de uma MLP, qual a diferença entre a atualização instantânea e a por lotes?**

**R-** Na atualização instantânea, calcula-se e aplica-se a correção a a cada par apresentado.

Já na atualização por lote, calcula-se os para todos os pares, e aplica-se a correção a , depois que todos os pares foram apresentados.

Na atualização instantânea, a estimativa do gradiente não é muito boa, e os passos são dados de forma mais aleatória, enquanto que na atualização por lote temos uma aproximação mais precisa para o gradiente. Entretanto, na prática, o primeiro método apresenta resultados melhores (menor probabilidade de parar em mínimos locais).

9. Em que situações pode ocorrer *underfitting*? No que isso implica? Como isto pode ser resolvido?

R- O

10. Em que situações pode ocorrer *overfitting*? No que isso implica? Como isto pode ser resolvido?

R- O

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

R- O

12. Sabe-se que não existe uma regra ótima para determinar a topologia correta de uma rede neural. Entretanto, pode-se utilizar de métodos heurísticos para estabelecer sua configuração inicial. Considerando que a “quantidade de pesos sinápticos na rede neural deve ser igual ao número de amostras no conjunto de dados” e as definições a seguir, faça o que se pede.

* I é o número de entradas da rede neural
* O é o número de neurônios na camada de saída da rede neural
* T é a quantidade de amostras no conjunto de dados

a) Deduza a fórmula para calcular a quantidade de neurônios “adequada” para uma rede com uma única camada escondida.

b) Deduza a fórmula para calcular a quantidade de neurônios “adequada” para uma rede com duas camadas escondidas (com a mesma quantidade de neurônios em cada).

c) Calcule a quantidade de neurônios da(s) camada(s) escondida(s) utilizando as fórmulas obtidas em (a) e (b), considerando que I=16, O=2 e T=8192. Lembre-se que não existem quantidades fracionárias de neurônios.