EA072 – Inteligência Artificial em Aplicações Industriais (2s2020)
Exercícios Conceituais 1 – ECc1
Atividade Individual – Peso 8 – Data de Entrega: 20/11/2020

Questão 1)
No contexto de redes neurais artificiais, existem estratégias iterativas para ajuste de pesos sinápticos (como no caso de uma MLP) e existem estratégias em que os pesos sinápticos são ajustados numa única etapa, por exemplo, empregando quadrados mínimos (como no caso de uma rede ELM). O que justifica a diferença de metodologia para ajuste de pesos sinápticos? No caso de estratégias iterativas de ajuste de pesos, o que provoca a ocorrência de mínimos locais?

Questão 2)
Nas duas primeiras questões do ECp1, foi trabalhado o conceito de *ridge regression*, que resolve o seguinte problema de quadrados mínimos regularizados:

$$\min_{\mathbf{w}} \|X\mathbf{w}-\mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2,\ \lambda \geq 0 .$$

Uma outra estratégia de regularização é o LASSO. Assista ao vídeo no *link* a seguir, a partir do instante 1h17’48”, que trata dos conceitos de LASSO e de *Elastic Net*.

https://youtu.be/aP7D-JzYP18

(a) Apresente uma vantagem em potencial de LASSO sobre *ridge regression*. (b) Apresente uma vantagem e uma desvantagem de *Elastic net* sobre LASSO sozinho e sobre *ridge regression* sozinho. (c) Se você tivesse utilizado LASSO na Questão 1 do ECp1, qual teria sido o significado prático dos valores anulados pelo LASSO? (d) Se você tivesse utilizado LASSO na Questão 2 do ECp1, qual teria sido o significado prático dos valores anulados pelo LASSO?

Questão 3)
Dado que é sempre possível obter a variância de uma distribuição uniforme contínua:

https://proofwiki.org/wiki/Variance_of_Continuous_Uniform_Distribution

prove que $b=\sqrt{\dfrac{3}{n^{[q-1]}}}$ para se obter $\mathrm{Var}\left(x^{[q]}\right)=\mathrm{Var}\left(x^{[q-1]}\right)$ para a q-ésima camada de uma rede neural MLP com função de ativação tangente hiperbólica, pesos $W^{[q]}$ inicializados com uma distribuição uniforme $U\left[-b,+b\right]$, $x^{[q-1]}$ sendo o vetor de entrada e $x^{[q]}$ sendo o vetor de saída desta q-ésima camada. Para se chegar a esta prova, é preciso mostrar antes que:

$$\mathrm{Var}\left(x^{[q]}\right)=n^{[q-1]}\mathrm{Var}\left(W^{[q]}\right)\mathrm{Var}\left(x^{[q-1]}\right),$$

onde $n^{[q-1]}$ é o número de entradas da q-ésima camada, também denominado *fan-in*.
Para tanto, recorra ao texto de autoria de Wei Yang e intitulado “Initialization and Normalization in Deep Neural Networks”, fornecido como material de apoio.
<u>Observação</u>: Como estamos falando de inicialização dos pesos da rede neural, é esperado que a variância das entradas de qualquer camada seja baixa, evitando partir para o treinamento com neurônios já saturados e maximizando o valor das derivadas da função de ativação, que representa a condição mais favorável para o ajuste de pesos. Nessas condições, é possível aproximar a função de ativação tangente hiperbólica pela função identidade.

Questão 4)
Para um certo problema a ser resolvido explorando um espaço matemático de busca, cada solução candidata é um vetor de atributos, que corresponde a um ponto neste espaço de busca a ser explorado. Suponha que todos os atributos de cada solução candidata são valores em ponto flutuante restritos a um intervalo fechado. Sendo assim, para o atributo a_i, tem-se que $a_i \in [v_i^{[\min]}, v_i^{[\max]}]$. Valores dentro deste intervalo são denominados factíveis e valores fora deste intervalo são denominados infactíveis. **Proponha um operador de mutação** para este tipo de codificação tal que a sua aplicação sempre gere um valor factível para qualquer atributo a ser mutado. Procure também garantir que qualquer valor no intervalo fechado possa ser produzido pelo operador de mutação, embora valores no entorno do valor atual do atributo devam ter maior probabilidade de ocorrência.

Questão 5)
Num contexto de robótica evolutiva para navegação autônoma, em que as ações de controle são atualizadas em instantes regulares e discretos no tempo, deseja-se evoluir um controlador para o robô que execute a tarefa de navegação num certo intervalo de tempo pré-estabelecido. Os critérios de desempenho são: (1) Minimizar o número de colisões (*NC*); (2) Minimizar o número de mudanças de direção (*NMD*); e (3) Minimizar a distância até o alvo (*DA*). Sabendo que o ponto de chegada (alvo) é sempre o mesmo, mas existem *p* pontos de partida distintos para o robô (ou seja, o *fitness* do robô só fica definido após a execução de *p* tarefas de navegação independentes), **apresente uma formulação matemática para a função de *fitness*** a ser adotada, permitindo que o usuário do seu programa defina a importância relativa dos três objetivos a serem minimizados. Nota: Como a função de *fitness* deve sempre ser maximizada e espera-se que ela sempre seja positiva, se existe um objetivo (f_{obj}) a ser minimizado e ele pode se anular, então uma possibilidade para a função de *fitness* é:

$$f_{fitness} = \frac{1}{f_{obj} + 1}.$$

Questão 6)
Descreva o procedimento empregado pelo algoritmo para realizar a operação indicada:

(a) *Simulated annealing* para escapar de mínimos locais.
(b) ACO (otimização por colônia de formigas) para definir a probabilidade de uma formiga adotar uma dentre múltiplas decisões. Considere a existência de feromônio e de um termo heurístico.
(c) Estratégia evolutiva para busca em espaços contínuos.

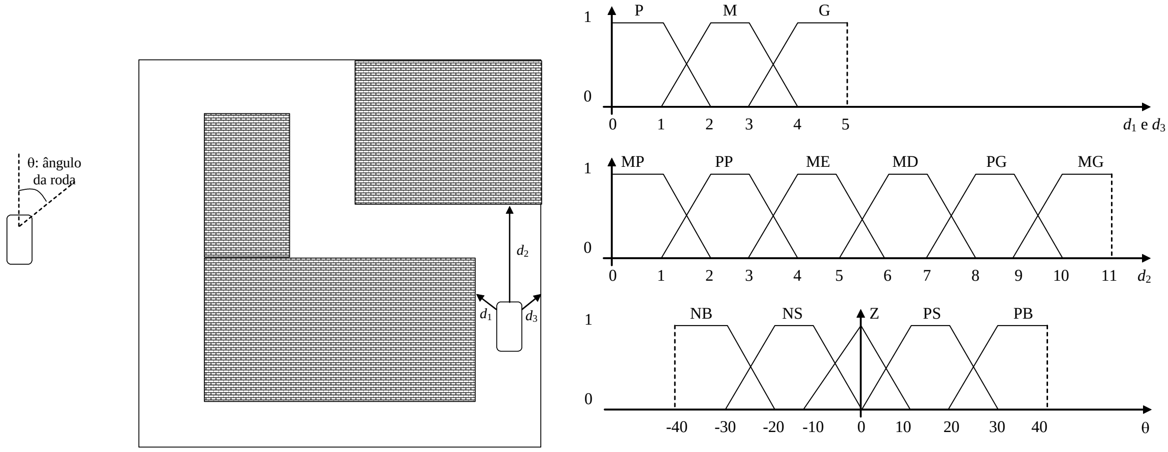
Questão 7)
Numa prova, um(a) estudante deve responder 7 questões de um total de 10 questões. **(a)** Quantas escolhas ele(a) tem? **(b)** Quantas escolhas ele(a) tem se entre as 7 questões a responder pelo menos 3 delas devem ser escolhidas dentre as 5 primeiras questões da prova?

Quantas coleções de *k* elementos existem, sabendo que existem *n* elementos candidatos?

	Repetição permitida	Repetição proibida
Ordenado	n^k	$(n)_k = \frac{n!}{(n-k)!}$
Não-ordenado	$\left(\binom{n}{k}\right) = \binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Questão 8)

Um veículo autônomo deve ser controlado por um sistema nebuloso, ou seja, um sistema baseado em regras nebulosas, e, para tanto, ele deve receber como entrada as informações de 3 sensores de distância (que produzem valores para d_1, d_2 e d_3) e fornecer na saída o ângulo da roda dianteira (θ), o qual é positivo no sentido horário.



As funções de pertinência associadas aos termos linguísticos que definem as partições das variáveis linguísticas d_1, d_2, d_3 e θ são fornecidas acima, à direita.

a) Proponha um consequente adequado e distinto para as regras ⟨SE d_1 é P E d_2 é PP E d_3 é G ENTÃO θ é ?⟩ e ⟨SE d_1 é P E d_2 é MP E d_3 é G ENTÃO θ é ?⟩. Justifique sua escolha.

b) Tomando {d_1 = 0,7; d_2 = 1,2; d_3 = 4,1} e supondo que as duas regras acima são as únicas ativas dentre todas as regras existentes, apresente o resultado de agregação dessas duas regras, assim como uma estimativa para o valor de θ.