

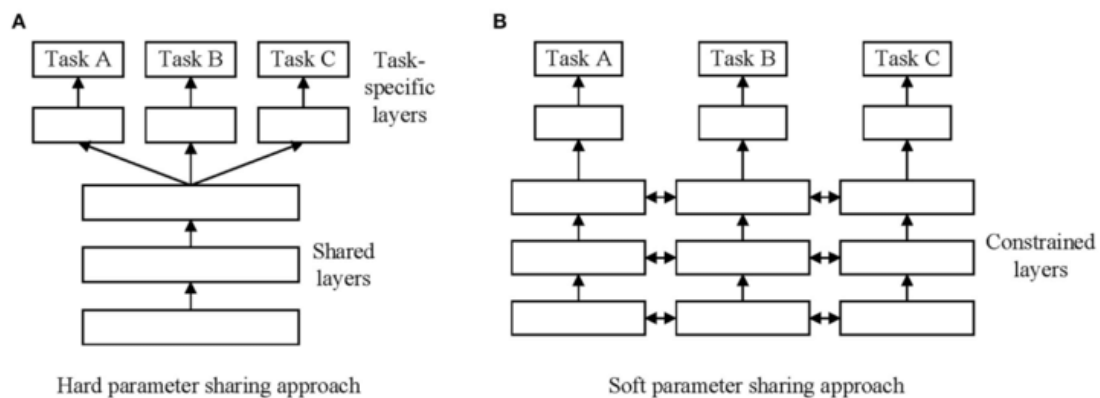
IA353 – Redes Neurais (1s2020)

Exercícios Conceituais 2 – EC 2 - Atividade Individual – Peso 3,5

Data de entrega da resolução (por e-mail): 04/08/2020

Questão 7) (1,5 pontos)

A figura a seguir, restrita a 3 tarefas de aprendizado, mas que poderia considerar qualquer número de tarefas, ilustra duas formulações possíveis para viabilizar a implementação de aprendizado multitarefa (MTL, do inglês *multi-task learning*) usando redes neurais artificiais. É fácil distinguir essas estruturas daquela associada ao aprendizado monotarefa (*single-task learning*), em que cada tarefa é aprendida de modo independente, cada qual empregando um modelo de aprendizado individual (basta retirar os fluxos bidirecionais horizontais da figura à direita, logo abaixo, e passam a existir 3 tarefas de aprendizado independentes).



Procure então responder:

- 7.1. O que é transferência negativa (*negative transfer*) e quando ela pode ocorrer em MTL?
- 7.2. Tomando a configuração da figura à esquerda, logo acima, explique como as camadas compartilhadas podem promover ganho de desempenho, quando se compara com a estratégia de 3 monotarefas independentes. Para tanto recorra a como deve se dar o ajuste de pesos dessas camadas compartilhadas, sabendo que há 3 tarefas de aprendizado envolvidas.
- 7.3. No link [<https://github.com/jiayuzhou/MALSAR/tree/master/manual>], baixe o Manual do MALSAR (*Multi-Task Learning via Structural Regularization*), associado a um pacote de soluções voltadas para modelos lineares de aprendizado. Em seguida, no contexto de MTL, explique quem é W e apresente uma conformação possível para os dois termos do problema de otimização da Eq. (1), página 6 do Manual do MALSAR, reproduzida a seguir, explicitando os objetivos práticos envolvidos na formulação escolhida por você.

$$\min_W \mathcal{L}(W) + \Omega(W)$$

Leituras de apoio em MTL:

<https://arxiv.org/abs/1707.08114>

<https://arxiv.org/pdf/1706.05098.pdf>

Questão 8) (1,0 pontos)

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

https://proofwiki.org/wiki/Variance_of_Continuous_Uniform_Distribution

prove que $b = \sqrt{\frac{3}{n^{[q-1]}}}$ para se obter $\text{Var}(x^{[q]}) = \text{Var}(x^{[q-1]})$ para a q -ésima camada

de uma rede neural MLP com função de ativação tangente hiperbólica, pesos $W^{[q]}$ inicializados com uma distribuição uniforme $U[-b, +b]$, $x^{[q-1]}$ 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:

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

onde $n^{[q-1]}$ é o número de entradas da q -ésima camada, também denominado *fan-in*.

Para tanto, recorra ao conteúdo em:

<https://eleg5491.github.io/initialization-and-normalization>

Observação: 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, o qual tende a ser elevado no início do processo de treinamento. Nessas condições, é possível aproximar a função de ativação tangente hiperbólica pela função identidade.

Questão 9) (0,5 pontos)

Ciente de que aprendizado de máquina envolve fundamentalmente aprender a partir de dados, e em sintonia com as tendências por maior *transparency & accountability* em tudo que permeia as técnicas de inteligência artificial, têm surgido iniciativas voltadas para uma padronização do processo de documentação de bases de dados disponíveis para treinamento. Sendo assim, consulte o material em:

<https://arxiv.org/pdf/1803.09010.pdf>

Em seguida, apresente as principais seções do padrão de documentação sugerido, assim como ao menos outras duas iniciativas recentes na literatura com propósitos similares.

Questão 10) (0,5 pontos)

Um dos papéis do estado-da-arte em qualquer área em franca expansão tecnológica é servir de referência para propostas ainda melhores, que então vão se tornar o estado-da-arte, numa evolução continuada. Não balizamos este curso de Redes Neurais visando atingir o estado-da-arte em *deep learning*, em qualquer frente que se queira considerar, seja pela imprevisibilidade da evolução tecnológica de ponta, seja pelo fato de o curso estar voltado para aspectos mais conceituais e introdutórios. Cabe, no entanto, uma última atividade conceitual voltada para algumas técnicas que conduzem a modelos de aprendizado estado-da-arte. Para tanto, procurem descrever os aspectos principais das metodologias que conduzem à *EfficientNet* e à *FixEfficientNet*, a partir dos seguintes textos da literatura:

<https://arxiv.org/pdf/1905.11946.pdf>

<https://arxiv.org/pdf/2003.08237.pdf>