

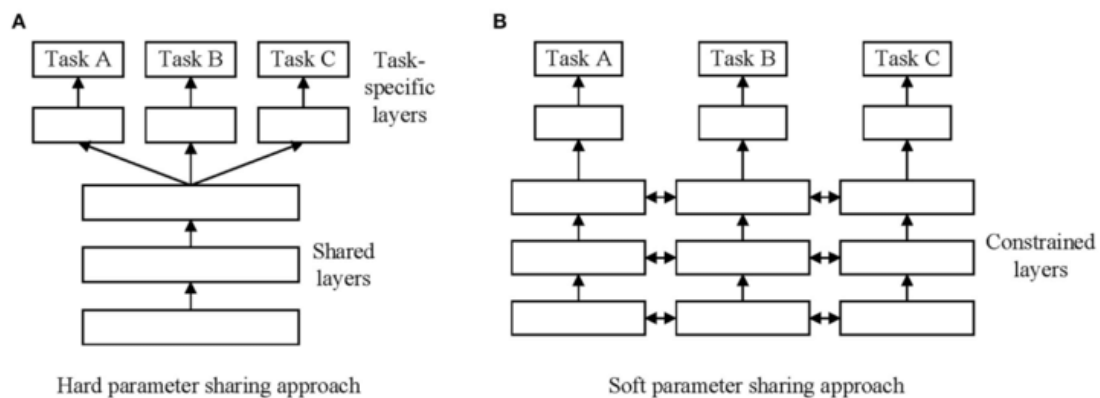
IA353 – Redes Neurais (1s2021)

Exercícios Conceituais 2 – EC 2 - Atividade Individual – Peso 5,0

Data de entrega da resolução: 22/06/2021 (Relatório Individual)

Questão 5) (1,8 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 6) (1,0 pontos)

Consulte os *links* a seguir para uma excursão por técnicas simples de inicialização de pesos de redes neurais:

[Weight Initialization for Deep Learning Neural Networks \(machinelearningmastery.com\)](https://machinelearningmastery.com/weight-initialization-for-deep-learning-neural-networks/)

[Weight Initialization in Neural Networks: A Journey From the Basics to Kaiming | by James Dellinger | Towards Data Science](#)

São investigados a seguir conceitos vinculados à inicialização de pesos denominada *Xavier*, voltada para a função de ativação tangente hiperbólica. 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: Com foco na 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 7) (1,2 pontos)

Suponha um modelo de aprendizado de alto desempenho e estrutura interna de elevada complexidade, por exemplo, uma rede neural profunda composta por muitos módulos internos. Suponha que esta rede neural profunda tenha sido treinada com sucesso. É pertinente indagar acerca da relevância individual e do papel efetivamente desempenhado por esses módulos internos para a manutenção do desempenho global desta máquina de aprendizado. Esse estudo é denominado análise por ablação (do inglês *ablation analysis*), sendo útil na busca de arquiteturas otimizadas de redes neurais profundas e para examinar a robustez de uma rede neural a danos estruturais, indispensável em aplicações críticas de segurança. Os estudos de ablação requerem que o modelo de aprendizado exiba degradação suave, de modo que ele continue a funcionar mesmo quando certos componentes estão ausentes ou têm seu efeito atenuado. A proposta da arquitetura GRU é um bom exemplo, no contexto de blocos LSTM, do sucesso da análise por ablação.

(7.1) Descreva uma técnica de ablação já proposta na literatura de redes neurais artificiais, indicando o contexto de aplicação e resultados obtidos.

(7.2) Consulte o paper a seguir e aponte as principais conclusões que a análise por ablação produziu ao se investigar a arquitetura VGG-19.

[\[1901.08644v2\] Ablation Studies in Artificial Neural Networks \(arxiv.org\)](#)

Questão 8) (1,0 pontos)

Explique o conceito de convoluções separáveis (do inglês *separable convolutions*) evidenciando seus tipos e suas principais vantagens sobre os operadores tradicionais de convolução. Sugestões de leitura inicial:

[Depth wise Separable Convolutional Neural Networks - GeeksforGeeks](#)

[A Basic Introduction to Separable Convolutions | by Chi-Feng Wang | Towards Data Science](#)

[Depth-wise Convolution and Depth-wise Separable Convolution | by Atul Pandey | Medium](#)