

# IA353A – Redes Neurais (FEEC/Unicamp)

Prof. Fernando J. Von Zuben

Notas do EC2

RA	Q5	Q6	Q7	Q8	Comentários
001658	1.8	1.0	1.2	1.0	
025809	1.5	0.9	1.2	1.0	Q5 (-0.3) No item 2, a sua formulação não é consistente, com a função perda dependendo de $Y^T$ (cuja definição também não está clara). O fundamental é mostrar como treinar as camadas compartilhadas com dados de todas as tarefas envolvidas e quando se ganha com isso. Q6 (-0.1) A tanh pode ser aproximada pela função identidade quando o seu argumento excursiona em torno da origem.
102388					Não foi entregue.
146244	1.6	1.0	0.7	0.85	Q5 (-0.2) No item 3, você não definiu a matriz $W$ como tendo os parâmetros de cada tarefa como colunas, sendo $T$ colunas em $W$ para $T$ tarefas. Q7 (-0.5) A análise está pobre, pois você apresentou basicamente uma tabela com resultados numéricos. Você não explicou como de fato ocorre a ablação de uma camada, importante para se entender cada linha da tabela apresentada. Q8 (-0.15) A explicação está tecnicamente consistente, mas faltaram estimativas de número de operações com as dimensões envolvidas, com e sem a separação.
156533	1.8	1.0	1.2	1.0	
164381	1.6	1.0	1.2	1.0	Q5 (-0.2) No item 3, você precisa detalhar melhor a notação. Você não definiu $W_i$ como sendo uma coluna de $W$ , de modo que $W$ tenha tantas colunas quanto tarefas e tantas linhas quanto atributos de entrada de todas as tarefas.
166812	1.5	1.0	1.2	0.85	Q5.1 (-0.3) Uma tarefa interferir no treinamento de outras é problema quando não há correlação entre essas tarefas, como você afirma mais adiante, em contraposição ao que afirmou antes. Portanto, não cabe esta afirmação isolada (não condicionada ao cenário em que ela é válida), pois MTL está fundamentado justamente nessa “interferência” conjunta. Por que há piora de desempenho de tarefas que não pertencem a um certo grupo de tarefas? É necessário sustentar sua afirmação. Q8 (-0.15) A explicação está tecnicamente consistente, mas faltaram estimativas de número de operações com as dimensões envolvidas, com e sem a separação.

168116	1.8	1.0	1.2	1.0	
176625	1.8	1.0	1.2	1.0	
182716	1.2	1.0	1.2	1.0	Q5.1 (-0.3) Não dá para entender quando se afirma que a transferência negativa é a interferência entre o conhecimento prévio e uma nova informação. Quem é este conhecimento prévio? Quem é esta nova informação? MTL não está restrito ao emprego de uma arquitetura conjunta. Q5.3 (-0.3) $W$ não é a matriz de pesos da rede. Que rede? São modelos lineares cujos parâmetros livres correspondem às colunas da matriz $W$ . A função custo precisa ser devidamente apresentada, pois em MTL ela envolve todas as tarefas.
203416	1.6	1.0	1.2	1.0	Q5.3 (-0.2) $W$ não é a matriz de pesos da rede. Que rede? São modelos lineares cujos parâmetros livres correspondem às colunas da matriz $W$ , na forma como você definiu a matriz $A$ na Q5.2. A função custo precisa ser devidamente apresentada, pois em MTL ela envolve todas as tarefas.
211516	1.6	1.0	1.2	1.0	Q5.3 (-0.2) Especificar melhor a matriz $W$ . São modelos lineares cujos parâmetros livres correspondem às colunas da matriz $W$ .
217002	1.8	1.0	1.2	1.0	
228101					Não foi entregue.
229060	1.3	1.0	1.2	1.0	Q5 (-0.5) Nos itens 5.1 e 5.2, você se concentrou demais em cenários de transfer learning, deixando de lado cenários de multitask learning.
229960	1.4	1.0	1.2	1.0	Q5.1 (-0.3) Embora intuitivo, faltou dizer por que a ausência de correlação entre as tarefas pode levar a uma transferência negativa em MTL. A transferência pode trazer benefícios, mas a transferência negativa nunca traz benefícios mesmo. Q5.3 (-0.1) A função custo precisa ser devidamente apresentada, pois em MTL ela envolve todas as tarefas.
232881	1.8	1.0	1.2	1.0	
234613	0.9	1.0	0.5	0.5	Q5.1 (-0.2) A sua explicação representa um dos casos de transferência negativa, pois ela pode ocorrer também entre tarefas de igual nível de dificuldade. Q5.2 (-0.3) Você misturou conceitos válidos para modelos de aprendizado lineares nos parâmetros com o cenário da questão, que envolve o treinamento de redes neurais com camadas compartilhadas. Você não disse como se dá o treinamento dessas camadas compartilhadas. Q5.3 (-0.4) O que você quis dizer ao afirmar que a variável $W$ representa o parâmetro? Ela é uma matriz de parâmetros e é necessário definir suas dimensões e o papel de cada coluna desta matriz. O que os termos em

					azul e vermelho representam? O que se ganha ao otimizar esses termos? Q7.1 (-0.3) Ablação é muito mais do que concluir de forma genérica que retirar unidades individuais de uma MLP reduzem seu desempenho. Q7.2 (-0.4) Há muitas outras análises interessantes que foram realizadas, por exemplo, associada à relevância das camadas. Q8 (-0.5) Faltou detalhar melhor como funcionam essas convoluções, recorrendo a figuras e apresentando as dimensões envolvidas. Não é verdade que o resultado é o mesmo, usando convolução comum ou separável profunda (como você denominou).
234617	1.7	0.9	1.2	1.0	Q5.2 (-0.1) Está incompleta a explicação de como se dá o treinamento das camadas compartilhadas. Q6 (-0.1) A sua equação entre Eq. (6) e Eq. (7) está equivocada, não podendo levar à Eq. (7). Q7 A VGG-19 não tem 19 camadas convolucionais.
234619	1.6	1.0	1.2	1.0	Q5.3 (-0.2) Você precisa detalhar melhor a notação. Você não definiu $W_t$ como sendo uma coluna de $W$ , de modo que $W$ tenha tantas colunas quanto tarefas e tantas linhas quanto atributos de entrada de todas as tarefas.
234633	1.3	1.0	0.8	0.5	Q5.2 (-0.2) Está incompleta a explicação de como se dá o treinamento das camadas compartilhadas. Q5.3 (-0.3) Você não descreveu a composição da matriz $W$ , a qual tem tantas colunas quanto tarefas e tantas linhas quanto atributos de entrada de todas as tarefas. Você não mostrou como a função de perda contempla todas as tarefas num único índice. Q7 (-0.4) Tanto no item 1 como no item 2, há mais análises que foram feitas pelos autores e conclusões a que se chegou. É necessário também indicar como é feita a “eliminação” de uma camada inteira da rede neural, no item 2. Q8 (-0.5) Geralmente não são geradas as mesmas saídas que na versão não-separável. Há também desvantagens (e não apenas vantagens) potenciais associadas à convolução separável, particularmente a espacial. Faltou apontar a existência de dois tipos de convoluções separáveis. Faltaram estimativas de número de operações com as dimensões envolvidas, com e sem a separação.
235624	1.8	1.0	1.2	1.0	
258499	1.8	1.0	1.2	1.0	
261152	1.8	1.0	1.2	1.0	
261663	1.2	1.0	0.9	0.75	Q5.2 (-0.3) Não foi apresentada uma explicação de como se dá o treinamento das camadas compartilhadas. Q5.3 (-0.3) Você não descreveu a composição da matriz $W$ , a qual tem tantas colunas quanto tarefas e tantas linhas quanto atributos de entrada de todas as tarefas. Você não mostrou como a função de perda contempla todas as tarefas num único índice. Q7.1 (-0.1) Faltou enfatizar a forma como são definidos os grupos de neurônios. Q7.2 (-0.2) Faltou evidenciar algumas análises, como a capacidade de recuperar o desempenho com apenas uma época de <i>recovery training</i> , após a ablação, e o efeito das taxas de ablação. Q8 (-0.25) Faltou apontar a existência de dois tipos de convoluções separáveis. Faltaram estimativas de número de

					operações com as dimensões envolvidas, com e sem a separação.
262249	1.5	0.9	1.05	0.75	Q5.2 (-0.15) A explicação de como se dá o treinamento das camadas compartilhadas está incompleta. Q5.3 (-0.15) Falta detalhar melhor qual é a composição da matriz $W$ , a qual tem tantas colunas quanto tarefas e tantas linhas quanto atributos de entrada de todas as tarefas. Q6 (-0.1) Faltou apresentar o passo em que a $\tanh$ pode ser aproximada pela função identidade quando o seu argumento excursiona em torno da origem. Faltou um somatório quando se chega ao produto de duas variâncias. Q7.1 (-0.15) Faltou descrever o procedimento da técnica de <i>ablation</i> OBD. Q8 (-0.25) Não é verdade que o resultado é o mesmo, usando convolução comum ou separável. Faltaram estimativas de número de operações com as dimensões envolvidas, com e sem a separação, para os dois tipos.
265568	1.8	1.0	1.2	1.0	
265945	1.8	1.0	1.2	1.0	