

Iniciado em	sábado, 29 Jan 2022, 15:16
Estado	Finalizada
Concluída em	sábado, 29 Jan 2022, 20:47
Tempo empregado	5 horas 31 minutos
Notas	6,50/8,00
Avaliar	8,13 de um máximo de 10,00(81%)

## Questão 1

Incorreto

Atingiu 0,00 de 1,00

Considere as seguintes equações para o GRU e para o LSTM:

**GRU: Gated Recurrent Unit**

$$\tilde{c}^{<t>} = \tanh(W_c[\Gamma_r * c^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_r = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>}$$

$$a^{<t>} = c^{<t>}$$

**LSTM: Long short-term memory**

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * \tanh(c^{<t>})$$

Qual das afirmativas abaixo sobre os gates da GRU é **incorreta**?

- ☐ a. Nenhuma das alternativas.
- ☐ b. O  $\Gamma_u$  da GRU tem papel análogo ao do update gate  $\Gamma_u$  da LSTM.
- ☐ c. O  $(1 - \Gamma_u)$  da GRU tem papel análogo ao do forget gate  $\Gamma_f$  da LSTM.
- ☒ d. Na GRU, a memória  $c^{<t-1>}$  é usada para processar a entrada e para propor o novo estado da memória.
- ☐ e. O  $a^{<t>}$  da LSTM pode estar fora do intervalo  $[-1, 1]$ .



Sua resposta está incorreta.

A resposta correta é:

O  $a^{<t>}$  da LSTM pode estar fora do intervalo  $[-1, 1]$ .

Questão 2

Correto

Atingiu 1,00 de 1,00

Considere uma GRU com dois neurônios na camada escondida, cuja entrada em cada tempo é um valor real (escalar). Por simplicidade, assuma que os bias são 0.

Dados  $x^{(1)} = 3$ ,  $\Gamma_r = [1.0, 0.5]^\top$ ,  $c^{(0)} = [1, 1]^\top$  e

$$W_c = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$$

$$W_u = \begin{bmatrix} 1 & 0 & 3 \\ -1 & 2 & -3 \end{bmatrix}$$

responda (usando pelo menos 3 casas decimais de precisão):

Quanto vale  $\Gamma_u$ ? (  ✓ ,  ✓ )<sup>T</sup>

Quanto vale  $\tilde{c}^{(1)}$ ? (  ✓ ,  ✓ )<sup>T</sup>

Quanto vale  $c^{(1)}$ ? (  ✓ ,  ✓ )<sup>T</sup>

Questão 3

Correto

Atingiu 1,00 de 1,00

$y^{\langle t-1 \rangle}$	$y^{\langle t \rangle}$	$\log \Pr(y^{\langle t \rangle}   y^{\langle t-1 \rangle})$	$y^{\langle t-1 \rangle}$	$y^{\langle t \rangle}$	$\log \Pr(y^{\langle t \rangle}   y^{\langle t-1 \rangle})$
$\emptyset$	o	-0.0	aplicações	$\langle \text{EOS} \rangle$	-0.0
o	aprendizado	-1.0	máquina	possui	-1.0
o	homem	-0.2	máquina	tem	-1.1
aprendizado	de	-1.0	muitas	aplicações	-0.1
aprendizado	profundo	-2.0	profundo	possui	-1.0
aprendizado	enobrece	-3.0	profundo	tem	-1.1
de	máquina	-0.9	possui	muitas	-1.0
enobrece	o	-0.1	tem	muitas	-0.5
homem	tem	-0.8	homem	possui	-0.6
homem	$\langle \text{EOS} \rangle$	-0.0			

Considere o modelo de linguagem neural parametrizado pela tabela acima, onde a probabilidade de uma sentença  $s = y^{\langle 1 \rangle}, \dots, y^{\langle T_y \rangle}$  é dada por  $\Pr(s) = \Pr(y^{\langle 1 \rangle}) \prod_{t=2}^{T_y} \Pr(y^{\langle t \rangle} | y^{\langle t-1 \rangle})$ . Note que neste modelo a saída no tempo  $t$  depende apenas da saída no tempo  $t - 1$  (ou seja, o estado oculto é completamente ignorado). Além disso, todas as transições que não aparecem na tabela tem probabilidade muito baixa.

Qual das frases seguintes tem maior probabilidade, segundo este modelo?

- ☐ a. O aprendizado profundo tem muitas aplicações.
- ☐ b. O aprendizado enobrece o homem.
- ☒ c. O aprendizado de máquina tem muitas aplicações.
- ☐ d. O aprendizado de máquina possui muitas aplicações.
- ☐ e. O aprendizado profundo possui muitas aplicações.



Sua resposta está correta.

As respostas corretas são:

- O aprendizado enobrece o homem.,
- O aprendizado de máquina tem muitas aplicações.

Questão 4

Correto

Atingiu 1,00 de 1,00

Qual das sequências a seguir seria gerada usando beam search com beam width  $B = 2$ ?

- ☐ a. O aprendizado de máquina possui muitas aplicações.
- ☐ b. O aprendizado enobrece o homem.
- ☒ c. O homem tem muitas aplicações.
- ☐ d. O aprendizado profundo tem muitas aplicações.
- ☐ e. O homem possui muitas aplicações.



Sua resposta está correta.

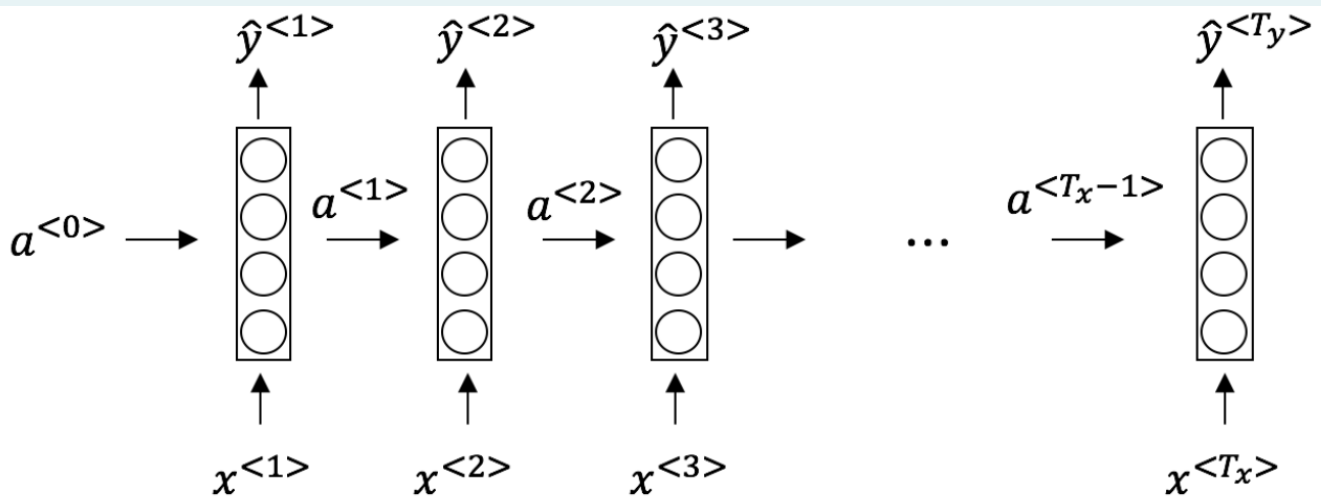
A resposta correta é:

- O homem tem muitas aplicações..

Questão 5

Correto

Atingiu 1,00 de 1,00



Qual das equações abaixo que relacionam as variáveis indicadas na Figura acima está **incorreta**?

- ☐ a.  $a^{(t)} = g(W_a[a^{(t-1)}, x^{(t)}] + b_a)$
- ☐ b.  $L(\hat{y}, y) = \sum_{t=1}^{T_y} L^{(t)}(\hat{y}^{(t)}, y^{(t)})$
- ☒ c.  $\hat{y}^{(t)} = g(W_{ya}a^{(t-1)} + b_y)$
- ☐ d.  $a^{(t)} = g(W_{aa}a^{(t-1)} + W_{ax}x^{(t)} + b_a)$
- ☐ e.  $\hat{y}^{(t)} = g(W_{ya}a^{(t)} + b_y)$



Sua resposta está correta.

A resposta correta é:

$$\hat{y}^{(t)} = g(W_{ya}a^{(t-1)} + b_y).$$

Questão 6

Parcialmente correto

Atingiu 0,50 de 1,00

Ainda com base na Figura anterior, suponha que cada entrada  $x^{(t)}$  e cada saída  $y^{(t)}$  seja uma palavra em *one hot encoding* de um vocabulário de tamanho 5000, e que o número de neurônios na camada oculta seja igual a 200. **Ignorando**-se os termos de bias, quantos parâmetros devem ser aprendidos? (Não use notação científica).

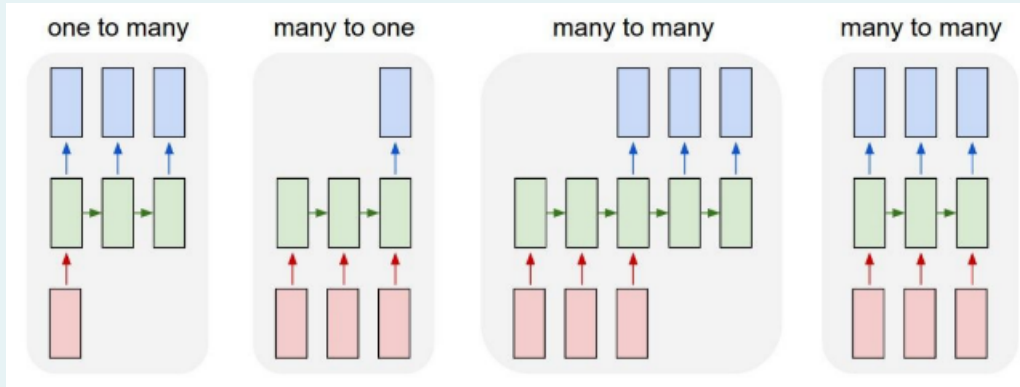
Resposta:  ☒

A resposta correta é: 2040000.

Questão 7

Correto

Atingiu 1,00 de 1,00



Da esquerda para a direita, as arquiteturas da figura acima podem ser usadas para:

- ☐ a. análise de sentimento, geração de texto, tradução automática, reconhecimento de entidades
- ☐ b. POS-tagging, identificação de autor, geração de texto, reconhecimento de fala
- ☐ c. geração de música, classificação de proficiência, geração de texto, análise de sentimento
- ☐ d. geração de receitas, identificação de autor, reconhecimento de entidades, reconhecimento de fala
- ☒ e. geração de receitas, identificação de autor, tradução automática, reconhecimento de entidades



Sua resposta está correta.

A resposta correta é:

geração de receitas, identificação de autor, tradução automática, reconhecimento de entidades.

Questão 8

Correto

Atingiu 1,00 de 1,00

Sobre word embeddings, podemos afirmar que:

- ☐ a. Podem ser treinados de maneira eficaz a partir de um corpus pequeno.
- ☐ b. São prontamente interpretáveis, desde que a dimensão do embedding não seja  $\gg 8$ .
- ☐ c. Nenhuma das alternativas.
- ☒ d. Permitem usar pré-treinamento em tarefas de NLP em que há poucos dados rotulados.
- ☐ e. São representações esparsas e, portanto, permitem a redução do número de parâmetros a serem aprendidos.



Sua resposta está correta.

A resposta correta é:

Permitem usar pré-treinamento em tarefas de NLP em que há poucos dados rotulados..

◀ Lista 02 de RNNs - com gabarito

Seguir para...



Prova 4 - Submeta seu rascunho aqui ▶

