

**Iniciado em** sábado, 7 nov 2020, 01:57

**Estado** Finalizada

**Concluída em** sábado, 7 nov 2020, 01:58

**Tempo  
empregado** 59 segundos

Questão 1

Completo

Vale 2,00 ponto(s).

Considere o uso de uma rede neural não recorrente (por exemplo uma rede densa) para o problema de, dado um caracter atual, prever os  $k$  próximos caracteres de uma sentença, sendo formulado como um problema de regressão. Por outro lado, podemos também utilizar arquiteturas recorrentes. Qual alternativa está correta na comparação da adequação das duas abordagens ao problema?

Escolha uma:

- ☐ a. As unidades densas capturam contexto e as relações de ordenação entre as instâncias, sendo mais adequadas para esse problema.
- ☐ b. Ambas são igualmente adequadas ao problema e devem produzir resultados similares, desde que os dados sejam preparados para cada caso
- ☒ c. Unidades recorrentes cumprem melhor os requisitos de prever sequências pois mantêm um sumário que está relacionado ao contexto de um caracter e sua ordem em uma sentença.
- ☐ d. As unidades recorrentes possuem memória que considera a ordem que as instâncias aparecem nas iterações. Redes com unidades densas assumem independência entre as instâncias podendo considerar os dados de entrada em ordem arbitrária, e portanto as densas serão sempre melhores para problemas como o de predição de sequências.

Questão 2

Completo

Vale 2,00 ponto(s).

Dentre as unidades recorrentes mais conhecidas, destacam-se a RNN tradicional, a GRU e a LSTM, em ordem da unidade com menor número de parâmetros, para aquela com maior número de parâmetros e portanto de maior capacidade. A GRU foi proposta mais recentemente do que as outras, ainda que tenha menos graus de liberdade (capacidade) do que a LSTM. Qual a principal motivação de usar GRU frente à LSTM?

Escolha uma:

- ☒ a. A GRU possui menos parâmetros, mas mantendo um mecanismo de adição de informação e esquecimento ao sumário, e por isso é geralmente mais fácil de otimizar do que a LSTM.
- ☐ b. A GRU é mais adequada apenas para problemas envolvendo sequência-para-sequência na estratégia encoder-decoder, nas demais tarefas a LSTM é sempre superior.
- ☐ c. A GRU produz resultados significativamente melhores na maior parte dos casos reais do que a LSTM, ainda que seja mais demorada para treinar por iteração.
- ☐ d. O mecanismo de esquecimento e atualização de novas informações nos sumários da LSTM se tornou obsoleto.

Questão **3**

Completo

Vale 2,00 ponto(s).

Escolha a alternativa que descreve a forma de implementação do mecanismo básico de atenção.

Escolha uma:

- ☐ a. O mecanismo é implementado por meio do alinhamento entre dois vetores,  $a$  e  $b$ , comumente aplicado por meio de um produto interno. Após esse processo utiliza-se um autoencoder com erro médio quadrático para aprender a atenção do vetor  $a$  em relação ao vetor  $b$
- ☒ b. Comparação da similaridade entre um vetor  $s$  e um conjunto de vetores  $h_i$ . Estas similaridades, depois de normalizadas, são utilizadas para realizar uma soma ponderada dos vetores  $h_i$ , resultando em um vetor  $c$  de contexto
- ☐ c. Utilizando uma convolução unidimensional nos vetores  $a$  e  $b$  individualmente, e depois somando-os
- ☐ d. Por meio de uma unidade recorrente, computa-se uma série de sumários  $h$  ao longo de diversas iterações. Esses sumários são então processados por uma camada densa que computa a atenção com base na entrada.

Questão **4**

Completo

Vale 2,00 ponto(s).

Carregue a base de dados starbucks.csv, conforme abaixo, com uma divisão hold-out utilizando os 85% exemplos iniciais para treinamento e os restantes para teste e normalize no intervalo 0-1.

Grafe a série temporal, e note que há uma modificação brusca na série relativa ao teste, próxima ao ponto 950, que torna difícil sua predição.

Crie duas redes recorrentes, uma baseada em LSTM e outra em GRU, para prever, ponto a ponto, a série temporal de testes. Elas devem conter ativações do tipo ReLU:

- Camada recorrente (LSTM/GRU) com 16 unidades
- Camada recorrente (LSTM/GRU) com 8 unidades
- Dropout de 0.2
- Camada densa de predição

Treine as duas redes com função de perda MSE, batch\_size = 1, por 15 épocas, utilizando Adam com taxa de aprendizado 0.001. Antes de instanciar cada modelo, compile e treine, inicialize as sementes com seed(2) e set\_seed(2).

Calcule o MSE de teste (considere arredondamento para 4 casas decimais) e trace o gráfico comparando os dados reais com os preditos pelas duas redes. O resultado obtido foi:

Escolha uma:

- ☒ a. LSTM com menor erro do que GRU, e apenas LSTM foi capaz de melhor prever o pico na parte final da série
- ☐ b. GRU com menor erro do que LSTM, mas apenas LSTM foi capaz de melhor prever o pico na parte final da série
- ☐ c. LSTM com menor erro do que GRU, mas apenas GRU foi capaz de melhor prever o pico na parte final da série
- ☐ d. GRU com menor erro do que LSTM, mas ambas falharam em prever o pico na parte final da série

Questão **5**

Completo

Vale 2,00 ponto(s).

Carregue o word embedding glove\_s50 em Português do NILC, conforme visto em aula, e obtenha os word embeddings das palavras: 'celular', 'triste', 'arte', 'livro', 'feliz', 'cansado', 'teclado', 'música', 'estudo', 'escritório'

Produza uma projeção em 2 dimensões utilizando o PCA com random\_state=1 e visualize os embeddings das palavras em um scatterplot.

Compute as distâncias (Euclidiana) entre "celular", "triste", "arte" e as outras palavras descritas acima, considerando a projeção obtida. Quais palavras (excluindo elas mesmas) estão mais próximas?

Escolha uma:

- ☐ a. celular-teclado, triste-feliz, arte-música
- ☒ b. celular-escritório, triste-cansado, arte-livro
- ☐ c. celular-teclado, triste-feliz, arte-celular
- ☐ d. celular-estudo, triste-cansado, arte-música

[◀ Exercícios - soluções](#)

Seguir para...



[Avaliação ▶](#)