

Guia de Estudos para a Prova 2

Índice Geral

1	Tópicos 6 e 7.....	2
2	Tópico 8 – Parte 1.....	2
3	Tópico 8 – Parte 2.....	6
4	Tópico 8 – Parte 3.....	8
5	Tópico 8 – Parte 4.....	10
6	Tópico 8 – Parte 5.....	13
7	Tópico 8 – Parte 6.....	16

1 Tópicos 6 e 7

Não é apresentado um roteiro de estudos para os Tópicos 6 e 7, os quais também compõem a matéria da Prova 2.

2 Tópico 8 – Parte 1

1. Compare as funções de ativação logística, tanh, ReLu e Leaky ReLu.
2. Como funciona a função *softmax* aplicada a todas as saídas de um classificador (Nota: O número de saídas é igual ao número de classes)? Neste caso, qual é a interpretação que se dá a cada saída do classificador?
3. O que é informado por cada elemento do vetor gradiente, tão fundamental no treinamento de redes neurais artificiais?
4. Por que o treinamento supervisionado em redes neurais artificiais envolve passos iterativos de ajuste dos pesos sinápticos? O que deve ser definido a cada passo de ajuste?

5. O que são mínimos locais no treinamento supervisionado?
6. Qual a diferença entre ajuste de pesos padrão-a-padrão, em batelada e em *mini-batch*?
7. Dê exemplos de problemas que estão sendo resolvidos com sucesso, incluindo soluções tidas como estado-da-arte (nenhuma outra proposta da literatura conduz a resultados tão bons), ao se adotar a técnica de *deep learning*.
8. O que são hiperparâmetros em *deep learning* e por que é tão relevante definir valores adequados para eles?
9. Dada uma arquitetura de rede neural profunda e dados os hiperparâmetros do processo de treinamento, treinar a rede neural é ajustar o valor dos pesos sinápticos da rede. Supondo que você domina linguagens de programação usuais em *deep learning*, se a você é dada a tarefa 1 de definição da arquitetura e dos hiperparâmetros de uma rede neural profunda para um certo problema de regressão ou classificação, com dados já disponíveis para treinamento, além da

tarefa 2 de buscar o melhor desempenho possível, que procedimentos você adotaria para chegar até uma solução para essas tarefas 1 e 2? Repare que a questão não envolve apenas o treinamento da rede neural (ajuste dos pesos sinápticos), mas também a etapa que antecede o treinamento da rede neural. Ou seja, como você faria para definir a arquitetura da rede, os seus hiperparâmetros (não havendo nenhuma limitação sua para programar o que você precisa) e os pesos sinápticos? Suponha ao menos que os recursos computacionais são limitados, ou seja, não permitem o emprego de buscas exaustivas (aquelas que testam todas as possibilidades de solução-candidata).

10. Considerando a entrada da rede neural como uma imagem numa certa resolução, o que é a operação de convolução (na verdade, correlação cruzada) realizada em camadas convolucionais de redes neurais profundas?
11. Por que o resultado da aplicação de um filtro convolucional sobre uma imagem é uma outra imagem?

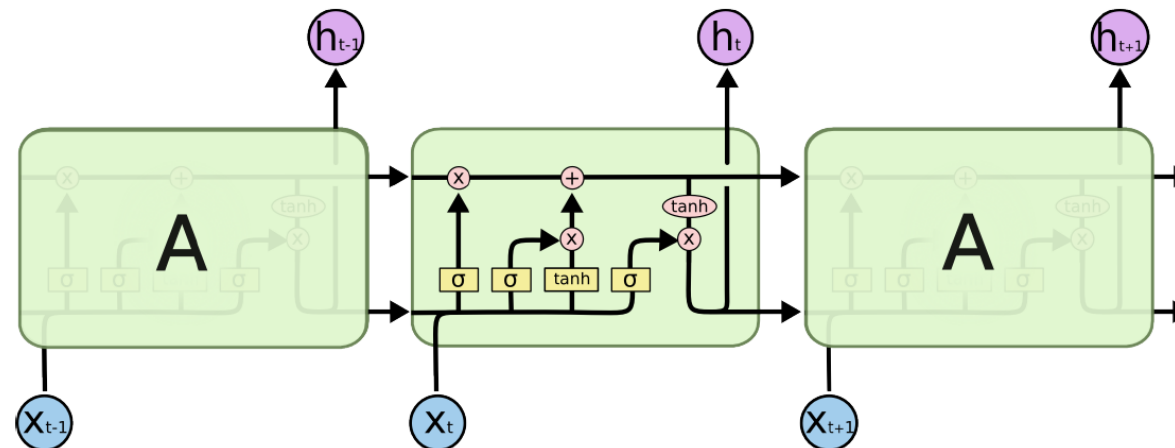
12. Apresente os benefícios de filtros de sub-amostragem.
13. Por que se diz que as camadas convolucionais, seguidas ou não por operações de sub-amostragem (e.g. *pooling*), são extratores de características?
14. O que é *padding* e o que é *striding* em uma convolução?
15. Calcule o número de pesos em uma certa camada de uma rede neural convolucional, dadas as dimensões envolvidas (imagem de entrada e hiperparâmetros).
16. Dadas algumas especificações de projeto, proponha dimensões adequadas para uma certa camada convolucional.
17. Qual é a relevância de estratégias de *transfer learning* em classificação de imagens, permitindo, por exemplo, ajustar apenas a última camada *fully connected* de acordo com as demandas específicas de cada aplicação?
18. O que é *dropout* e qual a sua principal função?
19. Como calcular os pesos em um treinamento empregando *dropout*?

20. Por que se diz que os componentes de um ensemble devem divergir no erro?
Apresente alguma técnica para gerar diversidade de comportamento entre modelos de aprendizado voltados para a solução de um mesmo problema.
21. O que são as operações de geração, seleção e combinação em ensembles?
22. Por que, mesmo havendo um componente que apresente um desempenho individual superior ao desempenho do ensemble, ainda assim é vantajoso adotar um ensemble?
23. Quais são as principais distinções entre um ensemble e uma mistura de especialistas?

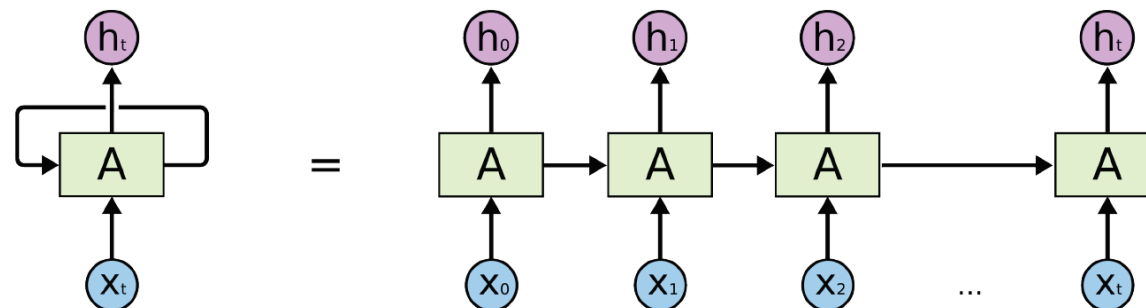
3 Tópico 8 – Parte 2

24. No contexto de sistemas dinâmicos, explique os conceitos de estado, dinâmica e trajetória no espaço de estados.
25. Compare as redes recorrentes tradicionais e os blocos LSTM.

26. Explique como operam as quatro estruturas internas (módulos em amarelo na figura a seguir).



27. Explique como funciona a estratégia denominada *backpropagation through time* (BPTT), usando a figura a seguir como apoio.



- 28. Com base na figura anterior, explique por que o bloco LSTM pode ser interpretado como uma rede neural profunda com pesos compartilhados.
- 29. Dê exemplos de aplicações práticas que requerem numa mesma máquina de aprendizado memórias de curto e longo prazos.

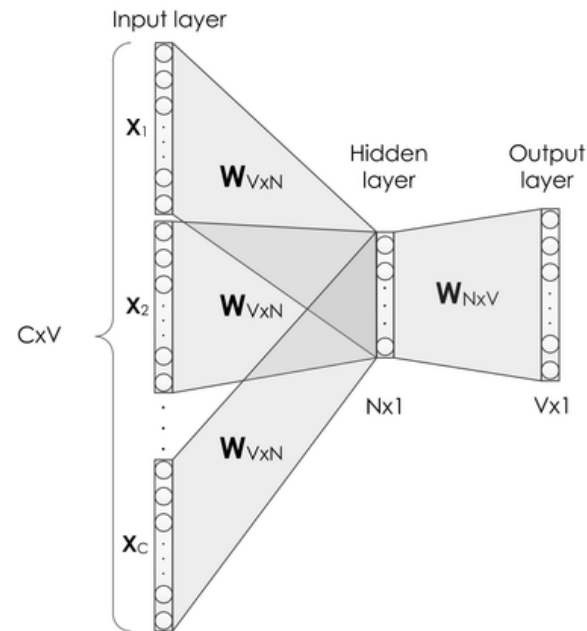
4 Tópico 8 – Parte 3

- 30. O que significa aprendizado da representação?
- 31. *Deep learning* geralmente envolve uma cascata de transformações não-lineares, ao longo das muitas camadas da rede neural. O que isto tem a ver com o aprendizado da representação?
- 32. Explique o conceito de variedade ou *manifold* em aprendizado de máquina.
- 33. Associe PCA com o conceito de variedade.
- 34. O que é um autoencoder?
- 35. Como autoencoders podem ser empregados na síntese de *manifolds*?

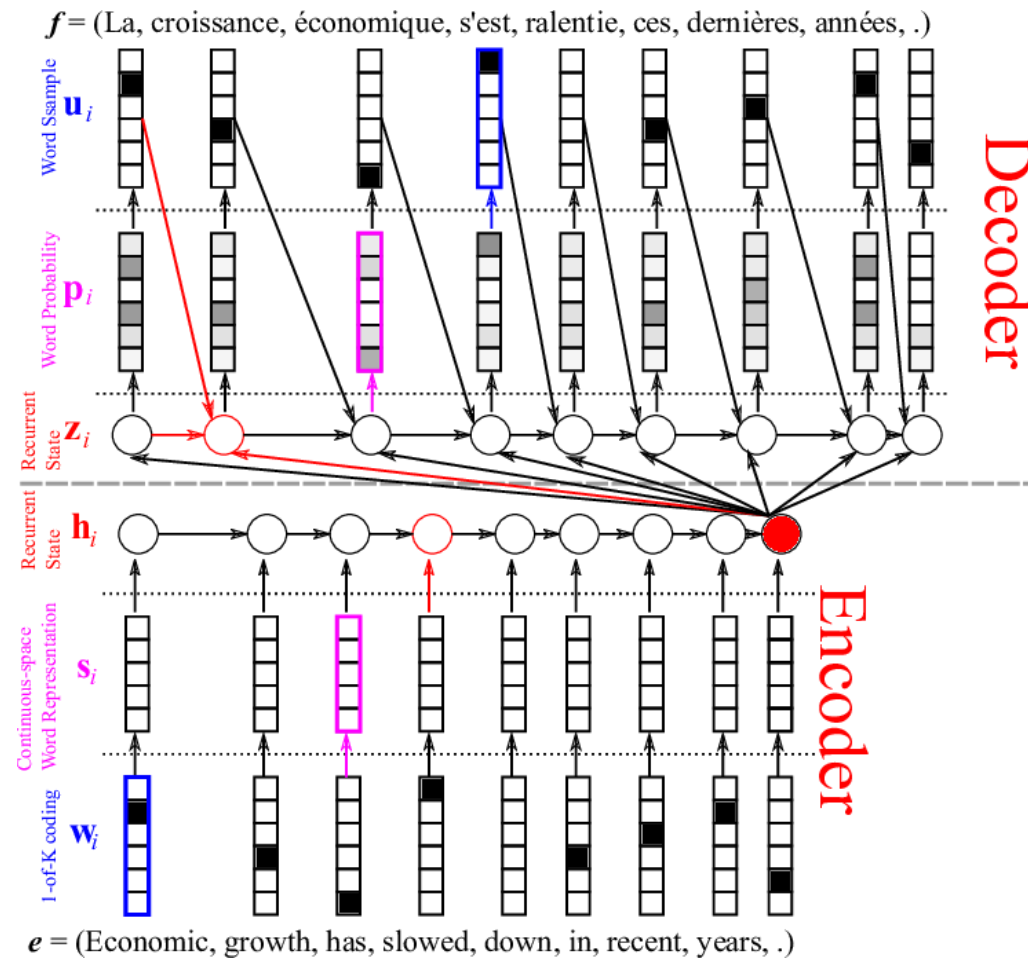
36. O que é um *denoising autoencoder*.
37. Como autoencoders podem ser utilizados na etapa de pré-treinamento de uma rede neural profunda? Usar aqui o conceito de *stacked autoencoders*.
38. Compare autoencoders tradicionais (AEs) e autoencoders variacionais (VAEs).
39. Como funciona o truque da reparmetrização, utilizado para viabilizar a retropropagação do erro pela camada de amostragem?
40. O que ocorre com as variáveis latentes na técnica de *disentangled VAEs*, supondo uma aplicação envolvendo imagens de faces humanas?
41. Apresente o princípio de operação de uma máquina de Boltzmann restrita (RBM).
42. Como RBM pode ser usada no pré-treinamento de uma rede neural profunda?
43. Como GANs (Tópico 8, Parte 5) podem ser empregadas para implementar neural inpainting?

5 Tópico 8 – Parte 4

44. Em processamento de linguagem natural (PLN), comente acerca das diferenças entre a codificação *one-hot* e a codificação densa, a qual também é denominada *word embedding*).
45. Explique como a codificação densa pode incorporar semântica em sua representação. Lembre-se que a codificação densa resulta do processo de treinamento vinculado à execução de alguma tarefa de PLN, como predição de próxima palavra ou da palavra central em uma frase.
46. Explique como opera a técnica *Continuous Bag of Words* (CBOW) no *word2vec*, ilustrada na figura abaixo.



47. Explique cada módulo que compõe a estrutura codificador-decodificador do tradutor de frases entre duas linguagens, apresentado na figura a seguir.
48. Procure justificar por que a proposta codificador-decodificador da figura a seguir é denominada de máquina de tradução estatística.



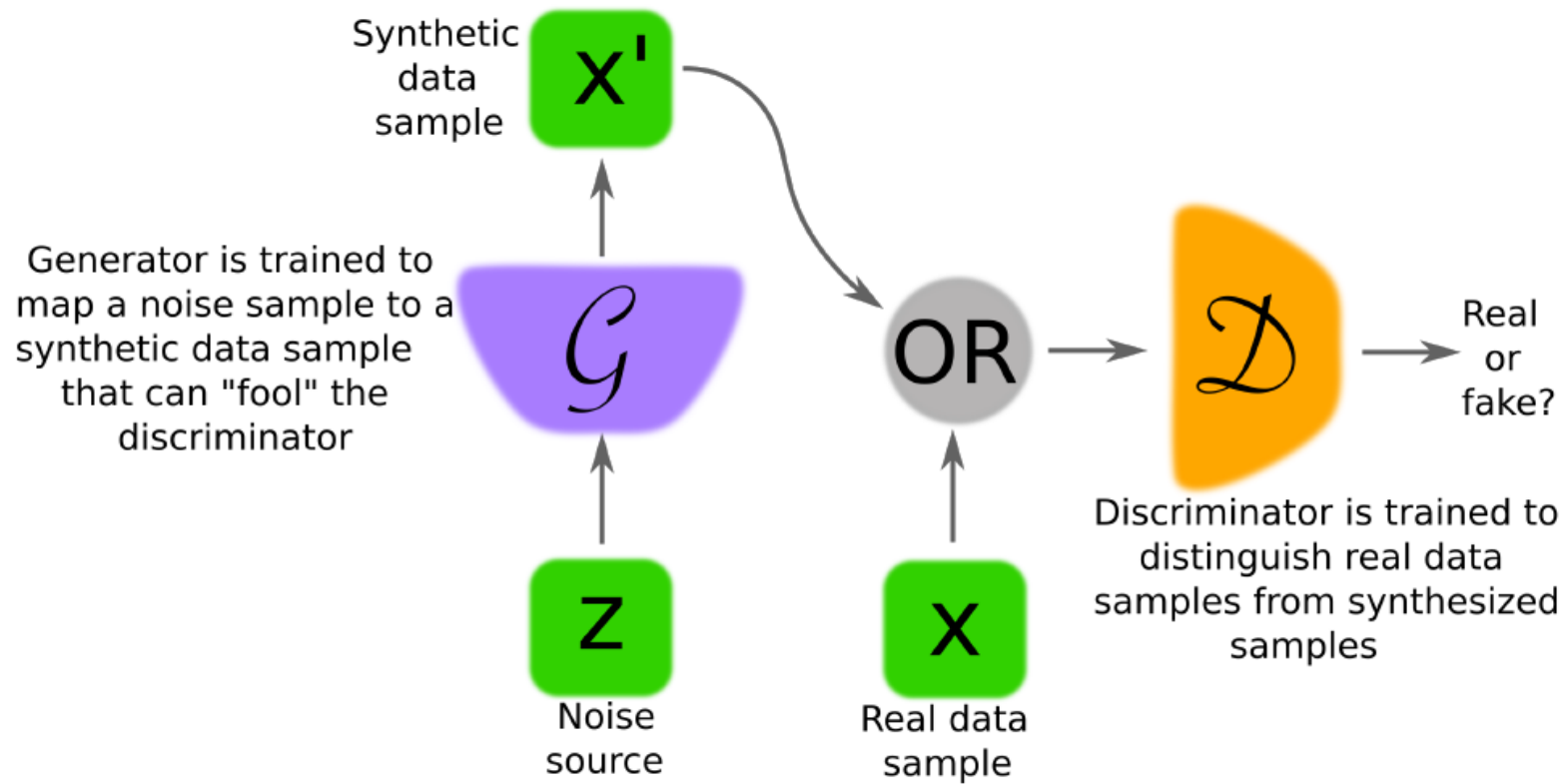
49. Qual a razão para se esperar um ganho de desempenho com o uso de um codificador bidirecional?

50. Explique como se dão os mecanismos de atenção em redes neurais profundas. Como eles podem auxiliar num processo de tradução de frases? Como eles podem auxiliar num processo de rotulação de uma imagem?
51. Por que é esperado que os *chatbots* envolvam habilidades adicionais, além do processamento de linguagem natural?

6 Tópico 8 – Parte 5

52. Procure justificar o crescimento acentuado no interesse por interpretabilidade em aprendizado de máquina, mais especificamente, em como as máquinas realizam suas inferências e predições.
53. Explique o conceito de *theory-guided data science*, o qual vem ganhando atenção em *data-intensive science*.
54. Explique o princípio de *activation maximization* ou *class prototypes*, explorado na parte 2 da Q1 do EFC3.

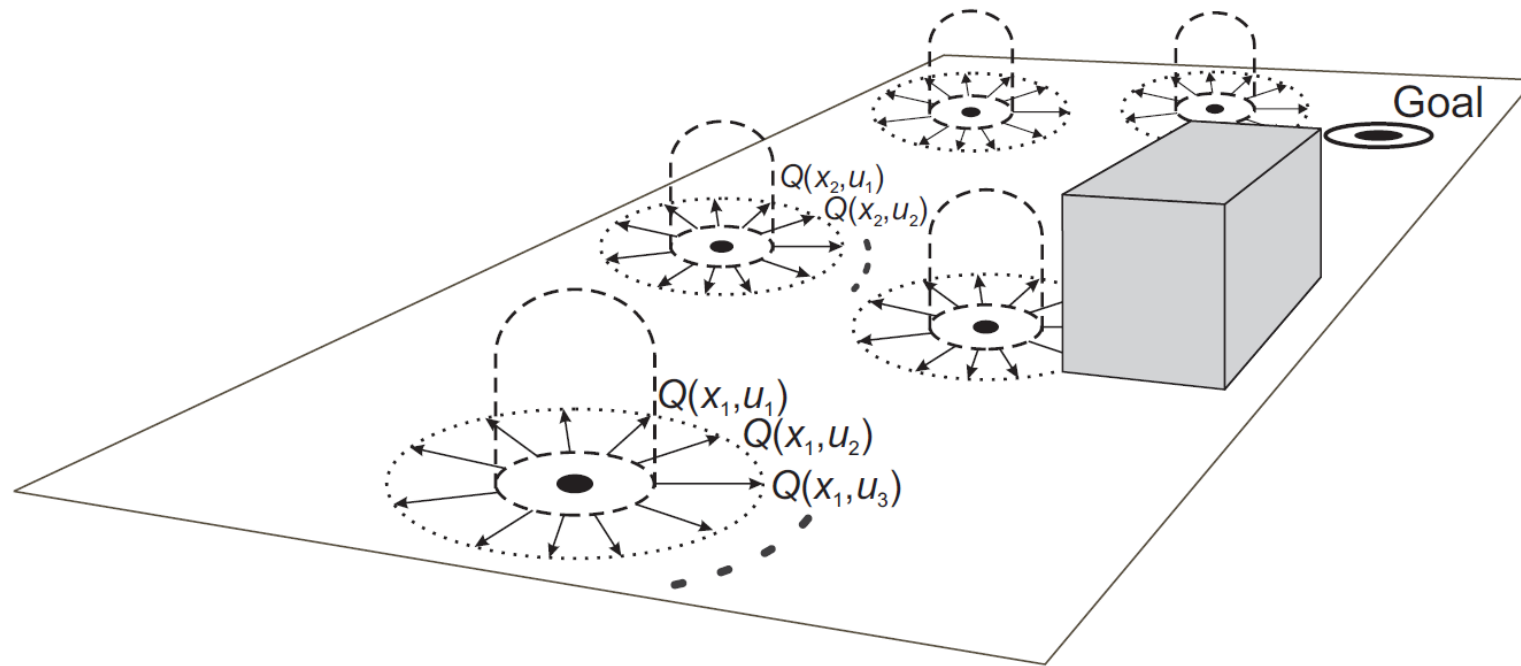
55. Compare conceitualmente análise de sensibilidade e *Layer-wise Relevance Propagation* (LRP).
56. Explique o conceito de mapa de calor em *Layer-wise Relevance Propagation* (LRP) e indique se há restrições na aplicação da técnica, conforme o tipo de máquina de aprendizado.
57. Como LRP pode ser usada na comparação entre máquinas de aprendizado?
58. Com base na figura abaixo, explique o princípio de operação do gerador e do discriminador, justificando a denominação de redes adversárias.



59. Como é possível guiar a geração de imagens usando um espaço de variáveis latentes *untangled*?
60. Apresente frentes de aplicação na literatura em que GANs têm se destacado.

7 Tópico 8 – Parte 6

61. Defina a técnica de aprendizado por reforço (não restrita a aprendizado profundo), enfatizando os conceitos de objetivo, estado, ação e recompensa.
62. Quais as principais diferenças entre aprendizado supervisionado e por reforço?
63. Explique o que é uma política e apresente o conceito de política ótima em tomada de decisão sequencial.
64. Num jogo de tabuleiro envolvendo dois jogadores, explique como uma rede neural pode aprender por reforço, jogando contra si mesma. Não é necessário recorrer aos detalhes técnicos de implementação.
65. Havendo recursos computacionais suficientes e supondo que o número de estados é finito e que o número de ações possíveis, estando em qualquer estado, é finito (veja uma ilustração a seguir, associada a uma tarefa de navegação de um robô móvel), indique como você faria para definir por busca exaustiva uma política de atuação ótima, independente do estado inicial do sistema.



66. Apresente um esboço de projeto de uma rede neural que vai aprender a jogar algum jogo de Atari, enfatizando as informações de entrada e saída.
67. Quais são os papéis do ator e do crítico no algoritmo *actor-critic*?
68. Quais são as diferenças entre uma tarefa de navegação autônoma de robôs e uma tarefa de controle de juntas de um robô móvel antropomórfico?