IA353 – Redes Neurais (1s2020)
Roteiro de Estudos (Parte 3)

Índice Geral

1 Tópico 8 (Parte 1) – Aspectos Adicionais de Otimização em Treinamento Supervisionado

1. No contexto de treinamento supervisionado em *deep learning*, apresente duas propostas independentes e mais elaboradas que o gradiente descendente com passo fixo (GDPF), que sejam capazes de, em média, produzir uma dada redução do erro de treinamento num menor número de épocas, quando comparado ao que se consegue com o GDPF.
2. Considere que a você é dada a missão de definir adequadamente o tamanho do *mini-batch* em treinamento supervisionado de uma rede neural profunda. Que estratégia você adotaria para cumprir esta missão?
3. Qual é o papel do termo de momentum?
4. Em que diferem as técnicas [SGD + *momentum*] e [*Nesterov accelerated gradient* (NAG)]?
5. O ADAM está entre os mais populares algoritmos adaptativos para o treinamento supervisionado em *deep learning*, sendo de fato uma proposta efetiva na redução

do erro de treinamento em um número reduzido de épocas. No entanto, sua eficiência está sendo questionada na literatura. Uma primeira limitação apontada foi a dificuldade de convergência na solução de um problema básico de otimização envolvendo quadrados mínimos. Qual foi a segunda potencial limitação do ADAM?

6. Explique o que é um método de otimização de 2a. ordem e procure justificar sua potencial relevância prática, mesmo sabendo que esse método é mais custoso, por iteração, que um método de otimização de 1a. ordem.
7. Seja para rede neural recorrente ou não-recorrente, o que são os fenômenos de gradiente explosivo e gradiente que se anula? Quais as consequências desses fenômenos para o treinamento supervisionado?
8. 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?

9. Apresente ao menos um motivo pelo qual a entropia cruzada se apresenta como uma função-custo mais adequada que o erro quadrático médio quando se treina classificadores na forma de redes neurais profundas.
10. Por que *batch normalization* tende a acelerar o processo de treinamento supervisionado em redes neurais profundas?
11. Consulte a literatura pertinente e procure entender como funciona a técnica *fixed-update initialization* (FIXUP), usada para inicializar os pesos na ResNet.
12. O que são hiperparâmetros em *deep learning* e por que é tão relevante definir valores adequados para eles?
13. 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).

2 Tópico 8 (Parte 2) – Redes Convolucionais + Dropout

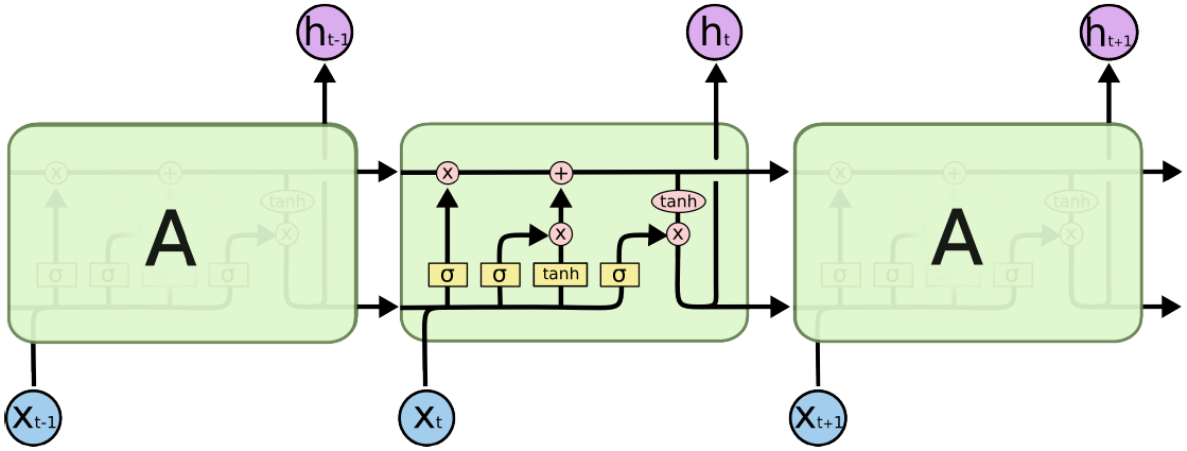
14. Considerando a entrada da rede neural como uma imagem numa certa resolução, o que é a operação de convolução (na verdade, correlação cruzada) realizada em camadas convolucionais de redes neurais profundas?

15. Por que o resultado da aplicação de um filtro convolucional sobre uma imagem é uma outra imagem?
16. Apresente os benefícios de filtros de sub-amostragem.
17. Por que se diz que as camadas convolucionais, seguidas ou não por operações de sub-amostragem (e.g. *pooling*), são extratores de características?
18. O que é *padding* e o que é *striding* em uma convolução?
19. Calcule o número de pesos em uma certa camada de uma rede neural convolucional, dadas as dimensões envolvidas (imagem de entrada e hiperparâmetros).
20. Dadas algumas especificações de projeto, proponha dimensões adequadas para uma certa camada convolucional.
21. 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?

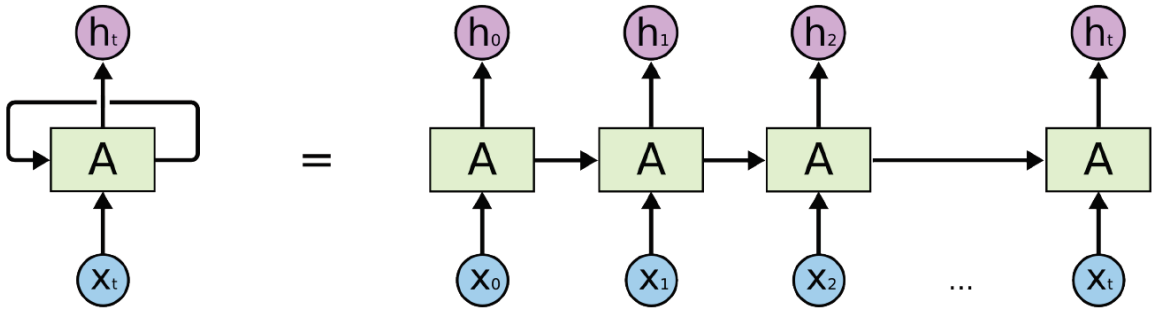
22. Descreva como opera o módulo básico da rede neural ResNet.
23. O que é *dropout* e qual a sua principal função?
24. Como calcular os pesos em um treinamento empregando *dropout*?
25. Por que se diz que os componentes de um ensemble devem divergir no erro? Apresente alguma técnica para gerar diversidade de comportamento entre modelos de aprendizado voltados para a solução de um mesmo problema.
26. O que são as operações de geração, seleção e combinação em ensembles?
27. Por que, mesmo havendo um componente que apresente um desempenho individual superior ao desempenho do ensemble, ainda assim é vantajoso adotar um ensemble?
28. Quais são as principais distinções entre um ensemble e uma mistura de especialistas?

3 Tópico 8 (Parte 3) – Bloco LSTM

29. No contexto de sistemas dinâmicos, explique os conceitos de estado, dinâmica e trajetória no espaço de estados.
30. Compare as redes recorrentes tradicionais e os blocos LSTM.
31. Explique como operam as quatro estruturas internas (módulos em amarelo na figura a seguir).



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



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

4 Tópico 8 (Parte 4) – Autoencoders, manifolds e RBMs

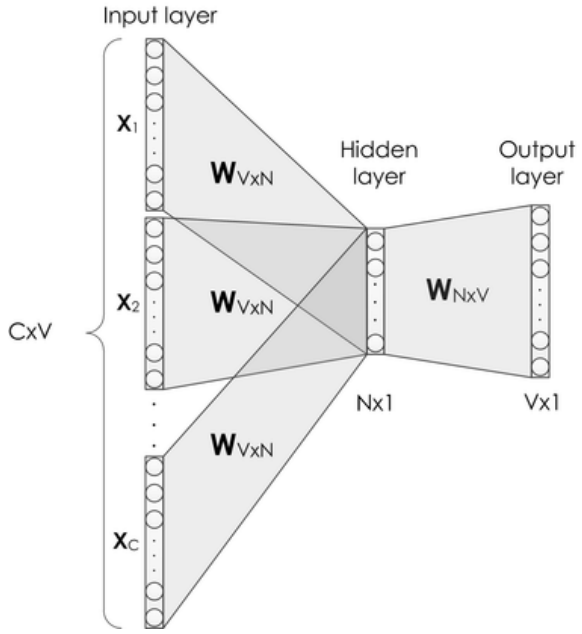
35. O que significa aprendizado da representação?
36. *Deep learning* geralmente envolve uma cascata de transformações não-lineares, ao longo das muitas camadas da rede neural. O que isto tem a ver com o aprendizado da representação?

37. Explique o conceito de variedade ou *manifold* em aprendizado de máquina.
38. Associe PCA com o conceito de variedade.
39. O que é um autoencoder?
40. Como autoencoders podem ser empregados na síntese de *manifolds*?
41. O que é um *denoising autoencoder*.
42. Como autoencoders podem ser utilizados na etapa de pré-treinamento de uma rede neural profunda? Usar aqui o conceito de *stacked autoencoders*.
43. Compare autoencoders tradicionais (AEs) e autoencoders variacionais (VAEs).
44. Como funciona o truque da repametrização, utilizado para viabilizar a retropropagação do erro pela camada de amostragem em VAEs?
45. O que ocorre com as variáveis latentes na técnica de *disentangled VAEs*, supondo uma aplicação envolvendo imagens de faces humanas?
46. Explique como opera a máquina de Boltzmann restrita (RBM).
47. Como RBM pode ser usada no pré-treinamento de uma rede neural profunda?

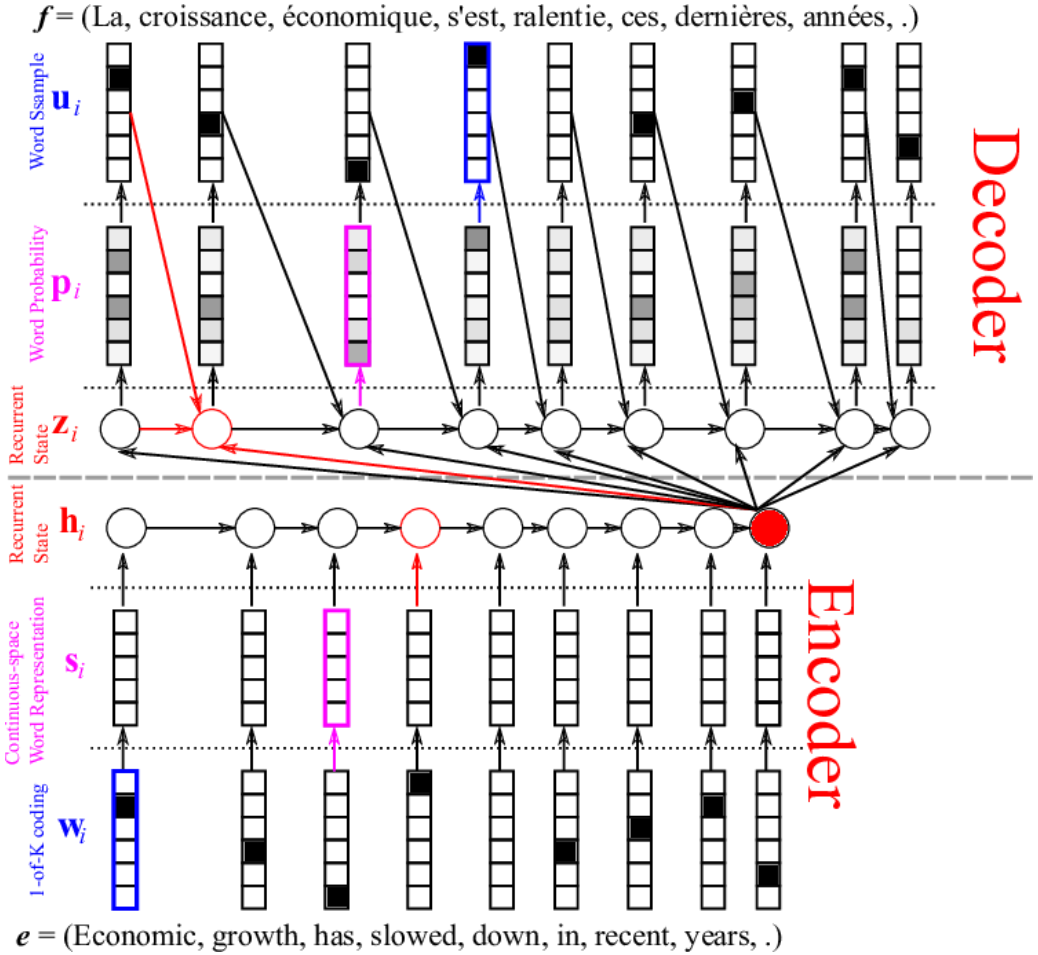
5 Tópico 8 (Parte 5) – Processamento de Linguagem Natural e Modelos de Atenção

48. Procure apontar impactos práticos de tecnologias que fazem com que computadores e seres humanos se comuniquem sem intermediários.
49. No contexto de linguagem natural, o que é um *corpus*?
50. Dê exemplos de mapeamentos vetor-para-vetor, vetor-para-sequência, sequência-para-vetor e sequência-para-sequência.
51. 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*).
52. 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.

53. Explique como opera a técnica *Continuous Bag of Words* (CBOW) no word2vec, ilustrada na figura abaixo.



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



56. O que é a técnica de *beam search* e como ela pode ser empregada para maximizar a chance de gerar a frase mais provável na saída?

57. Comente sobre índices de avaliação da qualidade da tradução realizada por uma máquina.
58. 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?
59. Qual a razão para se esperar um ganho de desempenho com o uso de um codificador bidirecional?
60. Por que é esperado que os *chatbots* envolvam habilidades adicionais, além do processamento de linguagem natural?
61. Consulte a literatura e procure entender como as técnicas de geração automática de rótulo para imagens (*image caption generation*) operam.
62. O que é o BERT e quais são suas duas estratégias de aprendizado?

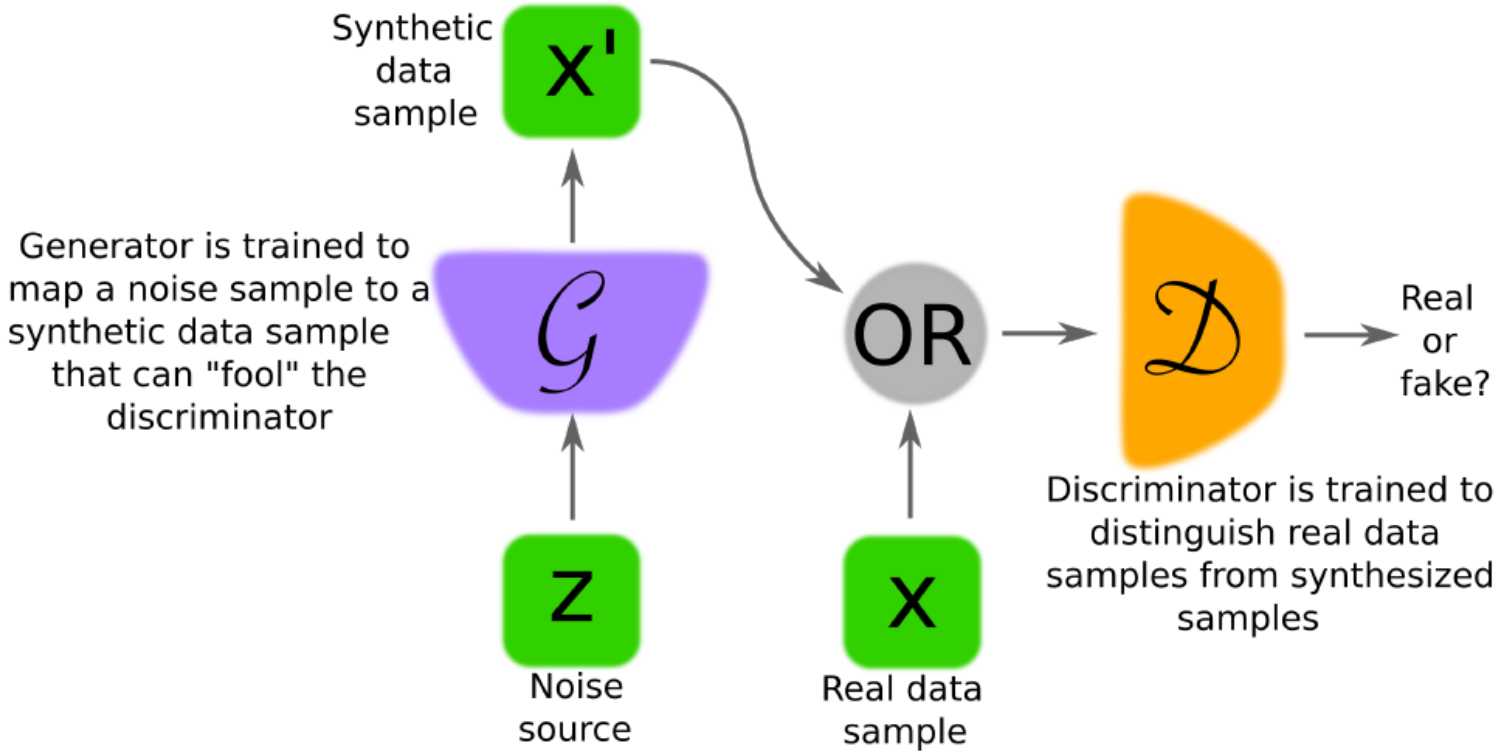
6 Tópico 8 (Parte 6) – Interpretabilidade em Redes Neurais Profundas

63. Procure justificar o crescimento acentuado no interesse por interpretabilidade em aprendizado de máquina, mais especificamente, em como as máquinas realizam suas inferências e predições.
64. Explique o conceito de *theory-guided data science*, o qual vem ganhando atenção em *data-intensive science*.
65. O nível de desempenho de um modelo de aprendizado é necessariamente conflitante com o nível de interpretabilidade deste mesmo modelo de aprendizado?
66. Como que o fato de se entender como um modelo de aprendizado funciona pode aumentar a confiança no (ou a aceitação do) próprio modelo de aprendizado?
67. Explique o princípio de *activation maximization* ou *class prototypes*, explorado no EFC3.

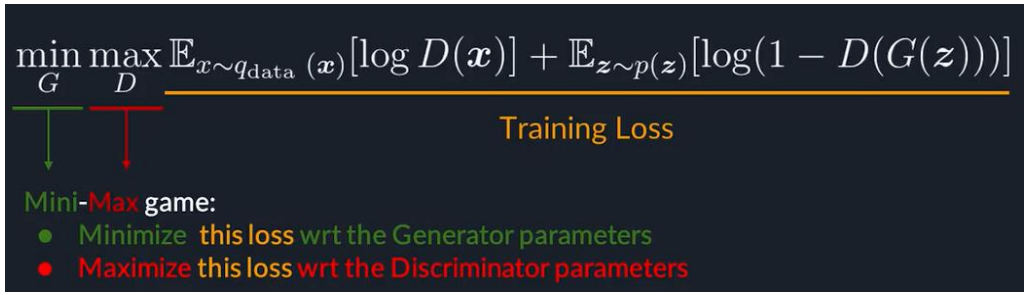
68. Compare conceitualmente análise de sensibilidade e *Layer-wise Relevance Propagation* (LRP).
69. Explique as propriedades de conservação, positividade, continuidade e seletividade exibidas pela técnica LRP.
70. O que você pode afirmar acerca da escalabilidade da técnica LRP?
71. Explique o conceito 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.
72. Como LRP pode ser usada na comparação entre máquinas de aprendizado?
73. Explique como realizar Spectral Relevance Analysis (SpRAy), usando LRP e spectral clustering.
74. Explique o que SpRAy fornece a respeito do comportamento de um classificador.

7 Tópico 8 (Parte 7) – Redes Neurais Adversárias Generativas

75. Com base na figura a seguir, explique o princípio de operação das redes neurais generativas adversárias. Procure justificar a vantagem em se considerar a atuação de uma rede discriminativa.



76. Explique conceitualmente a função objetivo de uma GAN, dada na forma:



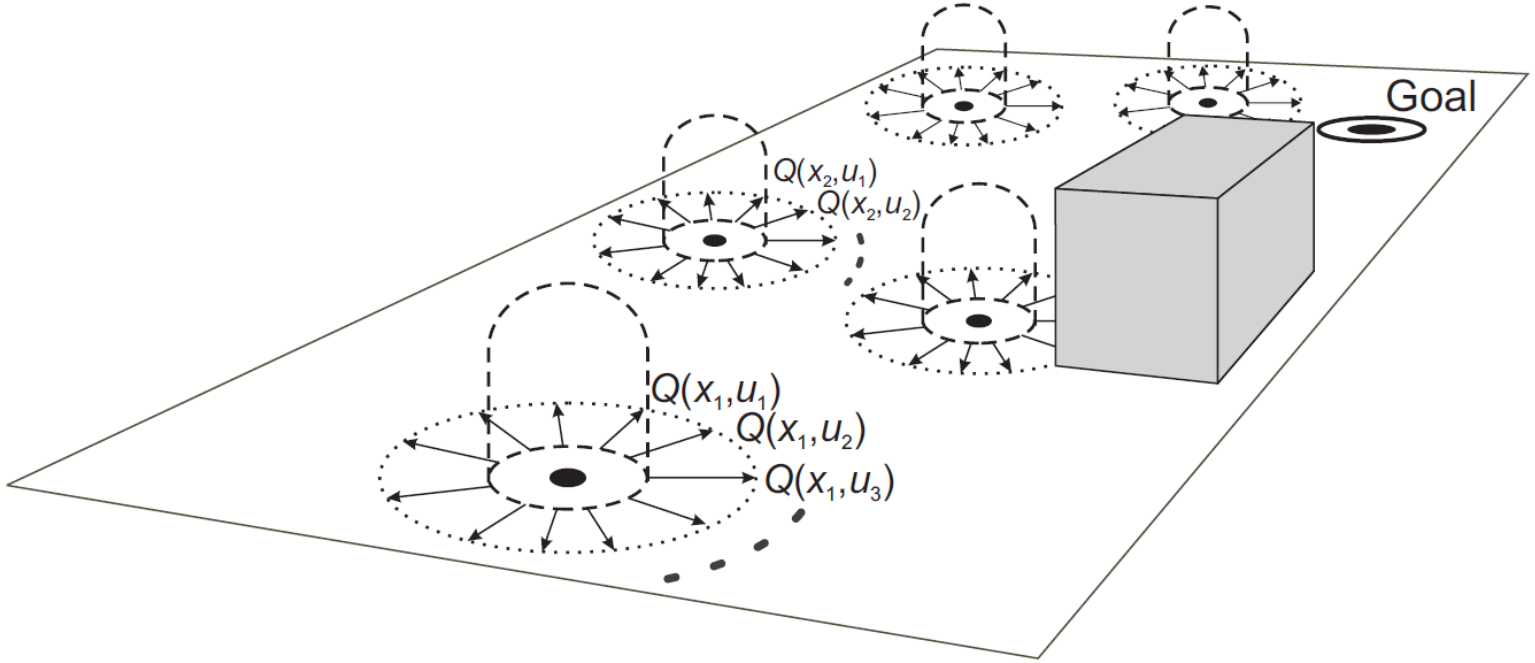
77. Qual é o equilíbrio que deve haver no *mini-max game*, conforme o treinamento progride?
78. O que é uma GAN construtiva e como ela é concebida?
79. Apresente os passos necessários para se operar com StyleGAN, dada uma rede neural generativa já treinada.
80. Como obter um ponto específico do espaço latente, a partir de uma *query image*?
81. Como definir as direções de caminhada no espaço latente para se manipular o estilo da face humana?

82. O que são *deep fakes*? Apresente ao menos uma técnica para se detectar *deep fakes*.
83. Como GANs podem ser empregadas para implementar *neural inpainting*, visto na Parte 4 deste Tópico 8?

8 Tópico 8 (Parte 8) – Introdução ao Aprendizado por Reforço

84. Defina a técnica de aprendizado por reforço (não restrita a aprendizado profundo), enfatizando os conceitos de objetivo, estado, ação e recompensa.
85. Quais as principais diferenças entre aprendizado supervisionado e por reforço?
86. Explique o que é uma política e apresente o conceito de política ótima em tomada de decisão sequencial.
87. 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.

88. 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.



89. 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.
90. Por que a solução produzida pelo *DeepMind AlphaStar* para o jogo *StarCraft II* é tão impressionante?
91. Apresente as diferenças entre os 3 tipos de aprendizado por reforço: model-based, value-based, policy-based.
92. Discorra acerca do dilema entre exploração e explotação e acerca dos desafios associados à atribuição de crédito em aprendizado por reforço.
93. O que é a função *Q*-valor?
94. O que é *Q-learning*?
95. O que é *Deep Q-learning*?
96. O que é gradiente de política?
97. Quais são os papéis do ator e do crítico no algoritmo *actor-critic*?
98. Quais são as diferenças entre uma tarefa de navegação autônoma de robôs e uma tarefa de controle de juntas de um robô móvel antropomórfico?