

Lista de Exercícios de Revisão: Inteligência Artificial e Aprendizado de Máquina

I. Fundamentos e Conceitos Iniciais

1. **Definição e Motivação do Aprendizado de Máquina:** Explique o que é o processo de aprendizado de máquina (ML) no contexto de um agente de *software*. Mencione dois motivos principais pelos quais não podemos simplesmente programar um agente corretamente desde o início, necessitando que a máquina aprenda.
2. **Indução e Dedução:** Diferencie os processos de **indução** e **dedução**. Por que as conclusões indutivas podem estar incorretas, enquanto as dedutivas são garantidas como corretas, *se as premissas forem corretas*?
3. **Tipos de Problemas de Saída:** Diferencie os problemas de **classificação** (quando a saída é um valor de um conjunto finito) e **regressão** (quando a saída é um número), dando exemplos de quando cada um é aplicado.
4. **Formas de Aprendizagem:** Descreva as três principais formas de aprendizado (supervisionado, não supervisionado e por reforço), detalhando o tipo de dados ou *feedback* que o agente recebe em cada caso.
5. **Aprendizado por Transferência:** Explique a abordagem de **aprendizado por transferência** e como ela difere do aprendizado que assume que o agente começa "do zero" a partir dos dados.

II. Aprendizado Supervisionado: Hipóteses e Avaliação

6. **Conceitos Fundamentais:** Dado um conjunto de treinamento de pares (x_i, y_i) gerados por uma função desconhecida $y = f(x)$, defina o que é a **função hipótese** (h) e o **espaço de hipóteses** (H).
7. **Conjuntos de Dados:** Qual é a verdadeira medida de uma hipótese? Descreva a função e o uso do **conjunto de teste** em comparação com o **conjunto de treinamento**.
8. **Viés (Bias) vs. Variância (Variance):** Explique o significado de **viés** e **variância** na avaliação de hipóteses. Como um modelo com **alto viés** (que tende ao subajuste) e um modelo com **alta variância** (que tende ao sobreajuste) manifestam erros diferentes nos dados?
9. **Overfitting e Underfitting:** Defina o que é **superajuste** (*overfitting*) e **subajuste** (*underfitting*). Qual é a relação entre esses conceitos e o *trade-off* viés-variância?
10. **Generalização:** O que significa dizer que uma hipótese (h) **generaliza bem** (*best-fit function*)?

III. Modelos Clássicos de ML (Árvores, Regressão e SVM)

11. **Árvores de Decisão (Estrutura):** Descreva a estrutura de uma **árvore de decisão**, mencionando o papel dos nós internos (teste de atributo) e dos nós-folha (valor de saída).
12. **Explique o que é *overfitting* em árvores de decisão** e cite duas formas de reduzi-lo (por exemplo, ajustando hiperparâmetros ou usando técnicas de poda).
13. **Estratégia Gulosa:** Qual é a estratégia utilizada pelo algoritmo de construção de árvores de decisão para escolher o melhor atributo a ser usado em cada etapa? O que caracteriza um **atributo forte** em oposição a um **atributo fraco** para a separação de classes?
14. **Entropia:** Defina **entropia** no contexto de aprendizado de máquina. Qual é a meta do algoritmo ao usar atributos para dividir o conjunto de dados em relação à entropia?
15. **Prevenção de Overfitting em Árvores:** Cite e explique duas estratégias comuns para prevenir o *overfitting* em árvores de decisão (que são um sintoma comum do problema quando são muito profundas).
16. **Regressão Linear (Conceito):** Qual é o objetivo principal da técnica de **Regressão Linear**?
17. **Regressão Linear Univariada:** Dada a forma da equação $y = w_1 \cdot x + w_0$, explique o significado dos termos **w_0 (Intercepto)** e **w_1 (Inclinação)**. Se você tiver um modelo ajustado $y = 2.5x + 10$, qual seria a previsão de y para $x=7$?
18. **Regressão Linear Multivariada:** Diferencie **regressão linear univariada** de **regressão linear multivariada**. No exemplo de previsão de preço de casa, se $w_1 = 2$ (área) e $w_2 = 25$ (número de quartos), o que esses coeficientes significam em termos práticos?
19. **Gradiente Descendente:** Explique o que é o **gradiente descendente** e qual é o seu papel como técnica de otimização (minimizando uma função de custo, como o Erro Quadrático) no treinamento de modelos.
20. **Tipos de Gradiente Descendente:** Compare as abordagens **Batch Gradient Descent**, **Stochastic Gradient Descent (SGD)** e **Mini-Batch Gradient Descent** em termos de custo computacional e estabilidade.
21. **Support Vector Machines (Margem):** Explique o que significa a **margem** em uma SVM e por que o objetivo da SVM é **maximizar essa distância**.
22. **Truque do Kernel (Kernel Trick):** O que é o **truque do kernel** e como ele permite que uma SVM crie uma fronteira de decisão não linear, apesar de o hiperplano separador ser linear? Cite e descreva a aplicação de três tipos comuns de kernels.
23. **Modelo Não Paramétrico:** Por que, especialmente quando usamos kernels não lineares, o SVM é frequentemente tratado como um modelo não paramétrico?

IV. Ensemble Learning e Regularização

24. **Ensemble Learning:** Qual é a ideia central do **Aprendizado em Conjunto (Ensemble Learning)** e como ele combina hipóteses individuais (modelos base) para formar um **modelo de conjunto**?
25. **Motivações para Ensembles:** Descreva como o aprendizado de conjunto pode levar à **redução de viés** (permitindo a representação de regiões complexas) e **redução de variância** (combinando classificadores descorrelacionados).
26. **Bagging:** Explique o funcionamento da técnica **Bagging (Bootstrap Aggregating)**, incluindo o uso da amostragem com reposição. Como ela é utilizada para reduzir a variância e prevenir o *overfitting*?
27. **Stacking vs. Bagging:** Qual é a principal diferença entre **Bagging** e **Stacking (Stacked Generalization)** em termos dos modelos base utilizados e dos dados de treinamento? No contexto do *stacking*, o que é um **meta-modelo**?
28. **Random Forests: Florestas Aleatórias (Random Forests)** são uma extensão do *bagging*. Além da amostragem com reposição dos dados (bootstrap), que outra amostragem aleatória é aplicada para construir árvores menos correlacionadas (subconjunto de atributos)? Por que isso é importante?
29. **Boosting:** Explique a abordagem sequencial do **Boosting**. Qual é o foco principal dessa técnica para melhorar o desempenho do modelo em iterações subsequentes?
30. **Gradient Boosting:** Como o **Gradient Boosting** se relaciona com o gradiente descendente? Cite duas estratégias de regularização que o Gradient Boosting utiliza para controlar o *overfitting*.
31. **Regularização (Conceito):** O que é **regularização** e qual é seu principal objetivo no treinamento de modelos de aprendizado de máquina?
32. **Função de Custo Regularizada:** Dada a função $\text{Custo}(h) = \text{EmpLoss}(h) + \lambda \times \text{Complexidade}(h)$, explique o que a **Perda Empírica (EmpLoss)** mede.
33. **Hiperparâmetro λ :** Explique o papel do hiperparâmetro **λ (lambda)**. Como o valor de λ (pequeno ou grande) influencia no risco de *overfitting* ou *underfitting*?

V. Redes Neurais Profundas (Deep Learning)

- 34.**História:** Quais pesquisadores publicaram o artigo que introduziu o primeiro modelo matemático de redes neurais em 1943? Qual limitação dos perceptrons simples (1958) causou o "Inverno da IA" em 1969?
- 35.**Retropropagação (Backpropagation):** Qual foi o papel do algoritmo de **retropropagação** (popularizado em 1986) no renascimento do interesse por redes neurais?
- 36.**Deep Learning:** Explique o que significa o termo "**Deep**" (Profundo) em *Deep Learning*. Compare a natureza dos caminhos de computação em um modelo raso (*shallow model*) versus um modelo de *Deep Learning* e como isso afeta a capacidade de representação de funções complexas.
- 37.**Redes Feedforward vs. Recorrentes:** Descreva a diferença fundamental na conectividade entre uma **rede feedforward** (conexões em apenas uma direção) e uma **rede recorrente (RNN)** (que possui ciclos e estado interno).
- 38.**Funções de Ativação:** Qual é o propósito da **função de ativação** em um nó (neurônio) de uma rede neural?
- 39.**Gradiente Desvanecido:** Explique o **problema do gradiente desvanecido** ao usar a função **Sigmoid**. Como isso afeta o treinamento de redes profundas?
- 40.**Função ReLU:** Por que a função **ReLU (Rectified Linear Unit)** se tornou popular na era moderna do *Deep Learning* (mencione o gradiente desvanecido e a esparsidade)?
- 41.**Função Tanh:** Como a função **Tanh (Tangente Hiperbólica)** se compara à Sigmoid em termos de simetria? Em qual tipo de problema, como a análise de sentimentos, o intervalo de saída de Tanh (centrado em zero) pode ser particularmente útil?

VI. Codificação e Arquitetura

- 42.**Codificação One-Hot:** Para atributos categóricos com mais de dois valores (ex.: tipos de cozinha), por que a codificação **one-hot** é preferível a mapear os valores para inteiros (e.g., 1, 2, 3, 4)?
- 43.**Função de Perda:** O que mede a **função de perda** (*Loss Function*) em uma rede neural?
- 44.**Camadas Ocultas:** Qual é o principal papel das **camadas ocultas** em uma rede neural profunda (DL)?

VII. Redes Convolucionais (CNNs)

- 45.**Desafios de Imagens em Redes Densas:** Explique o problema principal de tentar processar imagens grandes (ex.: 1 megapixel) usando uma rede totalmente conectada (densa) na primeira camada oculta, em termos de número de pesos necessários.
- 46.**Conexões Locais e Invariância:** Como as conexões locais resolvem o problema da explosão de parâmetros em CNNs, reduzindo drasticamente o número de pesos necessários em comparação com uma rede totalmente conectada? O que é **Invariância Espacial Local**?
- 47.**Kernels (Filtros):** O que são **kernels** (ou **filtros**) em redes convolucionais? Descreva como o *kernel* age como um detector de características (ex.: bordas).
- 48.**Convolução:** Explique o processo de **convolução**. O que é gerado ao deslizar o *kernel* sobre a entrada?
- 49.**Compartilhamento de Pesos:** Explique o conceito de **compartilhamento de pesos** nas CNNs.

VIII. Redes Recorrentes (RNNs e LSTMs)

50. **Arquitetura Recorrente:** Por que as **Redes Neurais Recorrentes (RNNs)** são adequadas para processar dados sequenciais (como séries temporais ou texto)?
51. **Atraso Temporal (\Delta):** Qual é o papel do **atraso temporal (Delta)** nas conexões recorrentes das RNNs?
52. **LSTM (Motivação e Portas):** Qual é o principal problema das RNNs básicas que as **LSTMs (*Long Short-Term Memory*)** foram projetadas para resolver? Descreva a função dos três *gates* (portas) principais da LSTM: **Forget Gate (f_t)**, **Input Gate (i_t)**, e **Output Gate (o_t)**.