

Inteligência Artificial

Aprendizado

Prof. Paulo Martins Engel

Computação por Aprendizagem

- Desde a invenção dos computadores têm sido feitos esforços para dotá-los com a capacidade de aprender.
- Em termos genéricos, aprender é melhorar o seu desempenho com a experiência.
- Aqui, experiência consiste em tentativas bem-sucedidas ou frustradas de resolver um problema.
- Apesar de ainda não sabermos como fazer os computadores aprender tão bem quanto as pessoas, hoje em dia existem vários algoritmos que são bem-sucedidos para certos tipos de tarefas de aprendizagem e foram feitos importantes avanços no entendimento teórico do aprendizado.

2

Aplicações bem-sucedidas de aprendizado de máquina

- Aprender a reconhecer palavras faladas: todos os sistemas de reconhecimento de fala empregam algum tipo de aprendizagem: redes neurais e modelos ocultos de Markov são as abordagens mais usadas.
- Aprender a guiar um veículo autonomamente: em 1989 o sistema ALVINN, baseado em RN, calculava o ângulo das rodas a partir de imagens de vídeo e de dados de um sensor de distância a laser e guiou um automóvel por 90 milhas, a 70 mph, numa auto-estrada.
- Aprender a classificar estruturas astronômicas: métodos de aprendizado de máquina têm sido usados pela NASA para aprender a classificar automaticamente objetos celestes, a partir de 3 Tb de dados de imagens.
- Aprender a jogar: vários programas de jogos bem-sucedidos têm sido desenvolvidos baseados em aprendizado: gamão (TD-Gammon), damas, etc.

3

Definição de um problema de aprendizagem

- Dizemos que um sistema computacional aprende da experiência E , em relação a uma classe de tarefas T e a uma medida de desempenho P , se seu desempenho nas tarefas em T , medido por P , melhora com a experiência E .
- Portanto, num problema de aprendizagem bem formulado identificamos 3 fatores: a classe das tarefas T , a medida de desempenho a ser melhorada P e a fonte de experiência (treinamento) E .
- Exemplo:
- **Problema de jogo de damas:**
- Tarefa T : jogar damas
- Medida de desempenho P : percentagem de jogos ganhos.
- Experiência de treinamento E : realizar jogos de damas contra ele mesmo.

4

Projeto de um sistema de aprendizagem

- Usando como exemplo um sistema para aprender a jogar damas.
- **Escolha da experiência de treinamento:**
- Se a experiência fornece informação *direta* ou *indireta* sobre as escolhas feitas pelo sistema.
- O sistema pode aprender *diretamente* de exemplos de configurações de tabuleiro e dos movimentos corretos respectivos.
- Ou o sistema pode aprender *indiretamente*, de seqüências de movimentos com os seus resultados finais. Neste caso, a informação sobre a correção de movimentos específicos durante o jogo deve ser inferida indiretamente do fato que o jogo foi vencido ou perdido.
- Aqui, o sistema se depara com o *problema de atribuição de crédito*.
- Outros fatores importantes são quanto ao controle da seqüência de jogadas (contra quem?) e da amplitude dos exemplos (fáceis, difíceis?).

5

Projeto de um sistema de aprendizagem

- **Escolha da função alvo:**
- Escolher qual o tipo de conhecimento que será aprendido.
- No jogo de damas, o problema pode ser colocado como um problema de *busca* pela *melhor* jogada entre as jogadas permitidas.
- A função alvo é a função heurística usada para guiar a busca. Ela atribui um escore numérico a uma dada configuração de tabuleiro.
- **Escolha da representação da função alvo:**
- A função adotada deve permitir alterar parâmetros durante o processo de treinamento: uma solução simples é adotar uma combinação linear de características do tabuleiro (número de peças pretas x_1 e brancas, x_2 , de damas pretas x_3 , de damas brancas x_4 , número de peças pretas x_5 e de peças brancas x_6 que podem ser capturadas na próxima rodada):

$$V(b) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

6

Projeto de um sistema de aprendizagem

- **Especificação dos exemplos de treinamento:**
- Exemplo de um estado em que as pretas ganham ($x_2 = 0$), $V(b) = +100$
 $\langle \langle x_1 = 3, x_2 = 0, x_3 = 1, x_4 = 0, x_5 = 0, x_6 = 0 \rangle, +100 \rangle$.
- **Regra para estimar valores de treinamento:**

$$V_{treino}(b) \leftarrow V(\text{Sucessor}(b))$$

- **Ajuste dos pesos:**
- Minimizar o erro médio sobre o arquivo de treinamento

$$E = \sum_{\langle b, V_{treino}(b) \rangle \in \text{exemplos de treinamento}} (V_{treino}(b) - V(b))^2$$

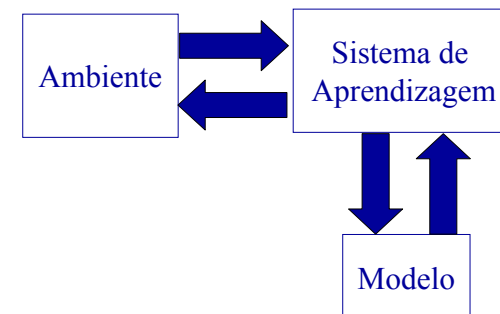
- **Regra de ajuste dos pesos LMS:** (descida de gradiente)

$$w_i \leftarrow w_i + \eta (V_{treino}(b) - V(b)) x_i$$

7

Sistemas de Aprendizagem

- Estamos interessados nos chamados Sistemas de Aprendizagem (SA), que são aqueles capazes de interagir com o seu meio-ambiente e, através dessa interação, extrair um modelo que pode ser utilizado para melhorar o seu desempenho.



8

Tipos de Aprendizado

- O aprendizado pode ser *supervisionado*, quando existe algum tipo de informação a priori (ou a posteriori) sobre as relações *causa-efeito* entre as ações do sistema (saída) e o estado do ambiente (entrada).
- Num sistema de aprendizado supervisionado por um *instrutor*, existe um *conjunto de treinamento* contendo pares de estados do ambiente e ações correspondentes, fornecidos pelo instrutor.
- Num sistema supervisionado por um *crítico*, existe apenas um sinal correspondente ao resultado, positivo ou negativo, de uma ação realizada (no passado).
- Além disso, o aprendizado pode ser *não-supervisionado*, se ele ocorrer apenas das observações dos estados do ambiente, sem conhecimento a priori.

9

Abordagens para a Construção de Modelos

- O modelo construído pelo processo de aprendizagem é limitado pela abordagem utilizada na sua representação.
- Nos sistemas de aprendizagem *simbólicos*, por exemplo, é dada ênfase na *descrição* do modelo, através de linguagens facilmente interpretadas por humanos.
- Nos sistemas de aprendizagem *computacionais*, na maioria das vezes, o interesse está centrado na *precisão* do modelo.
- Neste caso, do ponto de vista do usuário, o modelo é visto como uma caixa-preta.



10

Tipos de Tarefas de Aprendizagem

- Geração de modelos descritivos
 - Determinação de protótipos de agrupamentos
 - Determinação de co-ocorrências de valores de atributos
- Geração de modelos preditivos
 - Classificação
 - Regressão

11

Tipos de Problemas de Aprendizagem

- Os problemas estudados podem ser agrupados quanto à sua dependência temporal
 - Problema estático: só depende dos valores instantâneos das variáveis envolvidas
 - Problema dinâmico: além dos valores instantâneos, depende da sequência histórica das variáveis envolvidas

12

Modelos Descritivos

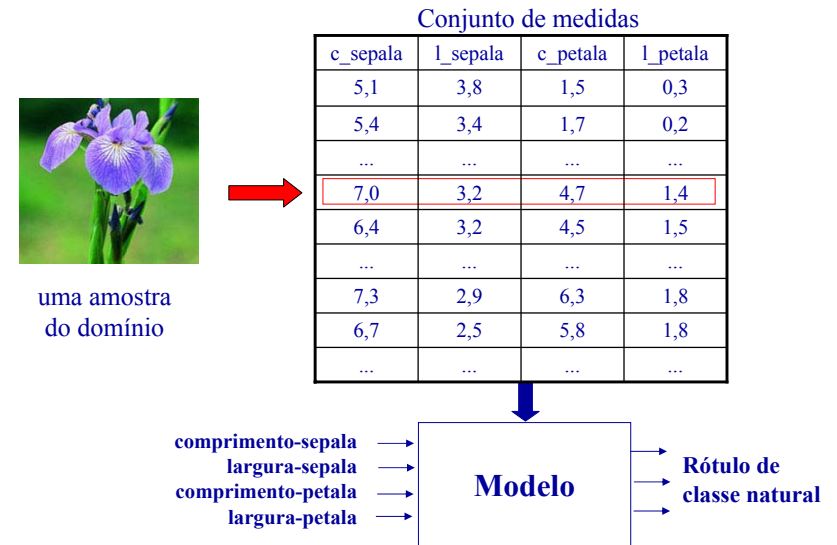
- A tarefa de geração de um modelo descritivo consiste (em grande parte) em analisar os dados do domínio (*entradas*) e sugerir uma *partição* do domínio, de acordo com *similaridades* observadas nos dados.



- Determinação de uma configuração de centróides de agrupamentos dos dados
- Modelo de agrupamento dos dados

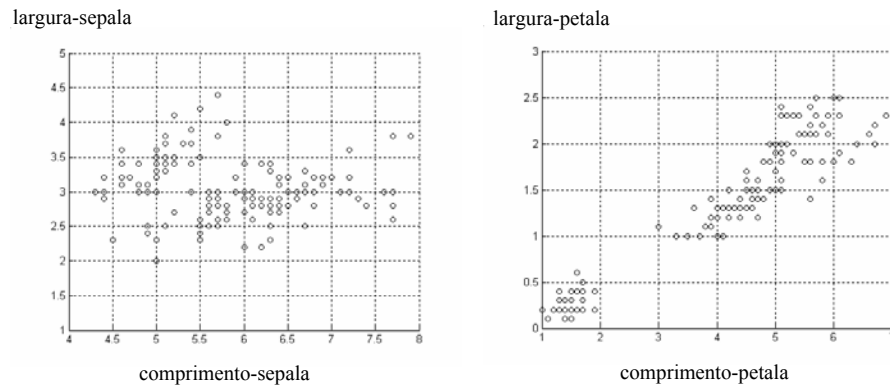
13

Dados para agrupamento de subespécies da planta Iris



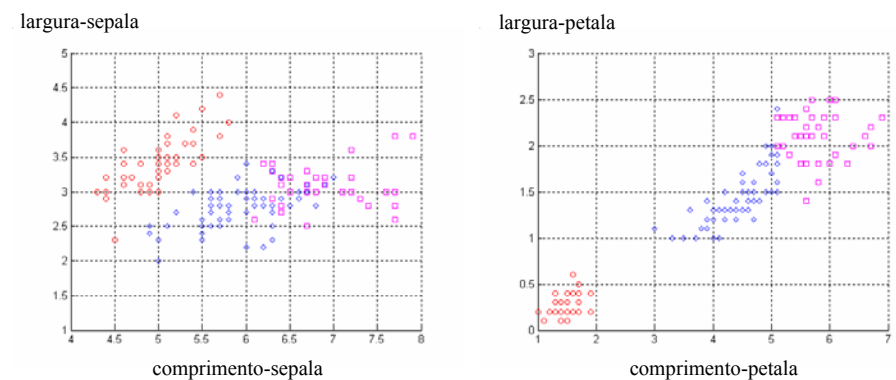
14

Dados da planta Iris



15

Modelo de agrupamento para os dados da planta Iris



16

Modelos Associativos

- Um modelo associativo é um caso especial de modelo descritivo.
- A tarefa de geração de um modelo associativo consiste em analisar os dados do domínio (*entradas*) e encontrar *co-ocorrências* de valores de atributos.
- Um modelo associativo é normalmente representado por um conjunto de *regras de associação*.



17

O problema da cesta de compras

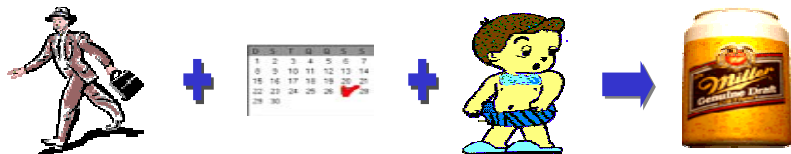
- O problema das cestas de compras (PCC) é o domínio típico para a tarefa de modelagem associativa.
- A tarefa consiste em analisar cestas de compras de clientes de um supermercado e determinar quais são os itens que freqüentemente são comprados juntos.
- A gerência do supermercado pode usar esta informação para desenvolver políticas de marketing e preço.

18

Aplicações e domínios

- Exemplo mais antigo de técnica de mineração de dados.
- Regra de associação:

Se cliente é *homem* e dia da semana é *sexta-feira* e compra *fralda*
Então compra *cerveja*



19

Modelos Preditivos

- A tarefa de geração de um modelo preditivo consiste em aprender um *mapeamento de entrada para a saída*.
- Neste caso, os dados contêm os valores de saída “desejados”, correspondentes para cada amostra.



- Classificação: saídas discretas representam rótulos de classe.
- Regressão: saídas contínuas representam valores de variáveis dependentes.

20

Dados para classificação de subespécies da planta Iris

amostras rotuladas
do domínio



Conjunto de medidas				rótulo
c_sepala	l_sepala	c_petala	l_petala	classe
5,1	3,8	1,5	0,3	Iris-setosa
5,4	3,4	1,7	0,2	Iris-setosa
...
7,0	3,2	4,7	1,4	Iris-versicolor
6,4	3,2	4,5	1,5	Iris-versicolor
...
7,3	2,9	6,3	1,8	Iris-virginica
6,7	2,5	5,8	1,8	Iris-virginica
...

comprimento-sepala
largura-sepala
comprimento-petala
largura-petala

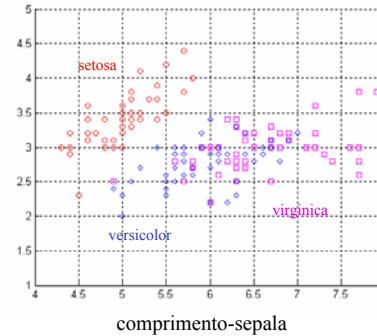
Modelo

Rótulo de classe
de interesse

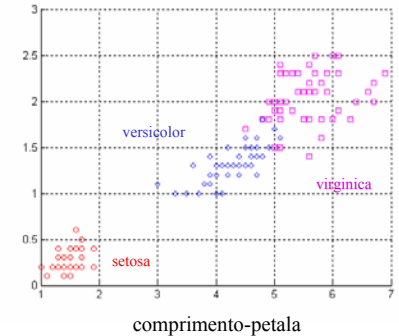
21

Dados para classificação de subespécies da planta Iris

largura-sepala



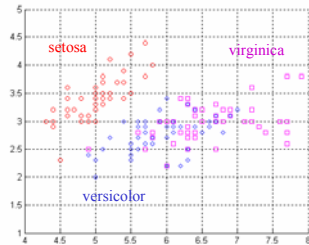
largura-petala



22

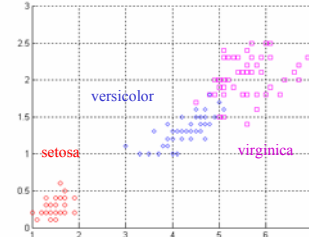
Classes de interesse versus classes naturais

largura-sepala

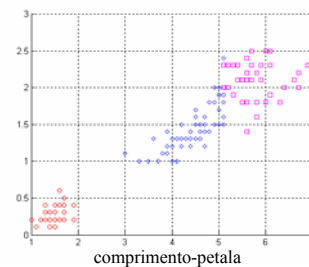
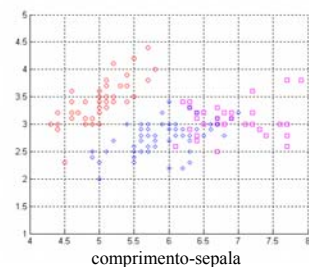


Classes
desejadas

largura-petala



Classes
naturais



23

Aprendizado por reforço

- O aprendizado por reforço é uma técnica que possibilita a aprendizagem a partir da interação com o ambiente.
- A interação com o ambiente nos permite inferir relações de causa e efeito sobre as consequências de nossas ações e sobre o que fazer para atingir nossos objetivos.
- Aprender por reforço significa aprender o que fazer - como realizar o mapeamento de situações em ações - de modo a maximizar um sinal numérico de *recompensa*.
- Não se dispõe da informação sobre quais ações devem ser tomadas, como é o caso no aprendizado supervisionado.
- O sistema de aprendizagem deve descobrir quais ações têm mais *chances* de produzir recompensa, e realizá-las.

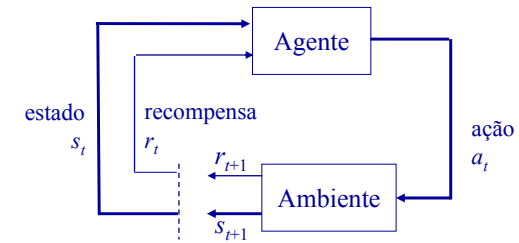
24

O problema da aprendizagem por reforço

- Os sistemas de aprendizagem e de tomada de decisão são denominados de *agente*.
- O objeto com o qual o agente interage é chamado de *ambiente*.
- A interação se dá com o agente selecionando ações e o ambiente respondendo a estas ações apresentando novas situações para o agente.
- O ambiente também fornece as recompensas - valores numéricos que o agente tenta maximizar ao longo do tempo.
- A cada instante t , o agente recebe uma representação do *estado* do ambiente $s_t \in S$, onde S é o conjunto de estados possíveis e, com isso, seleciona uma ação $a_t \in A(s_t)$, onde $A(s_t)$ é o conjunto de ações disponíveis no estado s_t .
- Num passo adiante de tempo ($t+1$), em parte como consequência da sua ação, o agente recebe uma recompensa numérica, $r_{t+1} \in R$, se encontrando no estado s_{t+1} .
- A cada instante, o agente implementa um mapeamento de estados em probabilidades de selecionar cada ação possível. Este mapeamento é chamado de política π_t , onde $\pi_t(s, a)$ é a probabilidade que $a_t = a$ e $s_t = s$.
- Pela aprendizagem, o agente muda a sua política, otimizando a recompensa.

25

O problema da aprendizagem por reforço



- Processo de decisão markoviano de tempo discreto finito:**
- O ambiente evolui probabilisticamente ocupando um conjunto finito de estados discretos ($s_t \in S$ é uma *variável aleatória*) formando uma *cadeia de Markov*.
- Para cada estado do ambiente há um conjunto de ações possíveis que podem ser realizadas pelo sistema de aprendizagem ($a_t \in A(s_t)$ é uma *variável aleatória*).
- Toda vez que o agente realiza uma ação, ele incorre em um certo custo ($r_{t+1} \in R$).
- A transição de estado ($s \rightarrow s'$) devido à ação a ocorre com uma certa probabilidade $\mathcal{P}_{ss'}^a$.

26

Componentes do aprendizado por reforço

- A tarefa é modelada de forma *discreta*, por um processo de decisão de Markov (PDM).
- O sistema de aprendizagem associa a cada estado s (discreto) do ambiente, um valor de utilidade $V(s)$ e mantém uma *memória* (distribuição) das ações tomadas em cada estado.
- A decisão de qual ação executar em cada estado é tomada com base numa *política* (π).
- A política corresponde a uma estratégia: escolher uma ação aleatória (*equiprovável*), ou escolher sempre a ação que correspondente ao maior valor de utilidade (*gulosa*), etc.
- Uma política acarreta uma *distribuição de probabilidades das ações* no PDM.
- A tarefa de aprendizagem consiste em encontrar a política *ótima* que maximiza o reforço.
- Considere um mundo de grade 4×4 , onde os estados 1 e 16 são terminais.
- Existem 4 ações possíveis (c, b, d, e : para cima, baixo, direita, esquerda).

Índices dos estados (s)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16



π equiprovável

V^π

	↖	↖	↖	
↖	↖	↖	↖	
↖	↖	↖	↖	
↖	↖	↖		

política gulosa π'

$V^{\pi'}$

	←	←	←	
↑	←	←	↓	
↑	→	→	↓	
→	→	→		

27

Aprendizado por Diferença Temporal

- É uma abordagem para o problema da *predição* (avaliação de política): estimar a função de valor V para uma dada política π .
- Para o outro problema, de controle, (encontrar uma política ótima) se usa alguma variante da iteração de política gulosa.
- TD (*Temporal Difference*) usa uma experiência adquirida seguindo a política π para calcular V_t , uma estimativa de V^π .
- Se um estado não-terminal s_t é visitado no tempo t , a sua estimativa $V(s_t)$ é atualizada baseado no que acontece após a visita.
- Lembrando que R_t é o retorno real, resultante do que segue o tempo t :

$$V^\pi(s) \leftarrow E_\pi \{R_t | s_t = s\}$$

- Uma forma iterativa de calcular a média de R_t é:

$$V(s_t) \leftarrow V(s_t) + \alpha [R_t - V(s_t)]$$

28

Algoritmo Q-Learning

- Analogamente a V , definimos o valor de realizar uma ação a no estado s seguindo uma política π , a *função de valor de ação para a política π* , $Q^\pi(s, a)$, como o retorno esperado quando, começando em s e realizando a ação a , seguimos a política π :

Inicialização: $Q(s, a)$, arbitrário

Repetir (para cada episódio):

 Inicializar s

 Repetir (para cada passo do episódio):

 Escolher a de s usando a política derivada de Q (p. ex. ϵ -gulosa)

 Realizar ação a , observar r, s'

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$$

$s \leftarrow s'; a \leftarrow a'$;

 até s ser terminal

O problema do penhasco

- Aprender a alcançar o objetivo, evitando cair no penhasco.

