

# AC2 - Inteligência Artificial

Eduardo Palhares Júnior

14 de outubro de 2020

## Parte I

# Introdução

1. Diz-se que um programa de computador aprende com a experiência  $E$  com relação a alguma tarefa  $T$  e alguma medida de desempenho  $P$ , se seu desempenho em  $T$ , medido por  $P$ , melhora com a experiência  $E$ . "Suponha que seu programa de e-mail observe quais e-mails você envia ou faz não marcar como spam, e com base nisso, aprende como filtrar melhor o spam. Qual é a tarefa  $T$  nesta configuração?"
  - a. Classifique e-mails como spam ou não spam.
  - b. Ajudar você a rotular e-mails como spam ou não spam.
  - c. O número (ou fração) de e-mails corretamente classificados como spam / não spam.
  - d. Nenhuma das outras opções, este não é um algoritmo de aprendizado de máquina.
2. Você dirige uma empresa e deseja desenvolver algoritmos de aprendizagem para resolver cada um dos dois problemas.
  - **Problema 1:** você tem um grande estoque de itens idênticos. Você deseja prever quantos desses itens serão vendidos nos próximos 3 meses.
  - **Problema 2:** você gostaria que o software examinasse contas de clientes individuais e, para cada conta, decida se ela foi hackeada / comprometida. Você deve tratar isso como classificação ou como problemas de regressão?
    - a. Trate ambos como problemas de classificação.
    - b. Trate o problema 1 como um problema de classificação, o problema 2 como um problema de regressão.
    - c. Trate o problema 1 como um problema de regressão, o problema 2 como um problema de classificação.
    - d. Trate ambos como problemas de regressão.

3. Dos exemplos a seguir, qual você trataria usando um algoritmo de aprendizado não supervisionado? (Marque todas as opções aplicáveis)
- a. Dado e-mail marcado como spam / não spam, aprenda um filtro de spam.
  - b. Dado um conjunto de artigos de notícias encontrados na web, agrupe-os em conjuntos de artigos sobre as mesmas histórias.
  - c. Dado um banco de dados de dados de clientes, descubra automaticamente segmentos de mercado e agrupe clientes em diferentes segmentos de mercado.
  - d. Dado um conjunto de dados de pacientes com diagnóstico de diabetes ou não, aprenda a classificar novos pacientes como portadores de diabetes ou não.

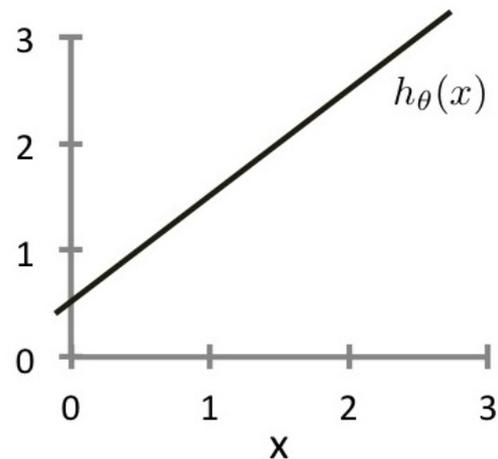
## Parte II

# Modelo e função custo

4. Considere o gráfico abaixo da função hipótese  $h_\theta(x) = \theta_0 + \theta_1 x$ .

Qual os valores de  $\theta_0$  e  $\theta_1$ ?

- a.  $\theta_0 = 0, \theta_1 = 1$
- b.  $\theta_0 = 0.5, \theta_1 = 1$
- c.  $\theta_0 = 1, \theta_1 = 0.5$
- d.  $\theta_0 = 1, \theta_1 = 1$

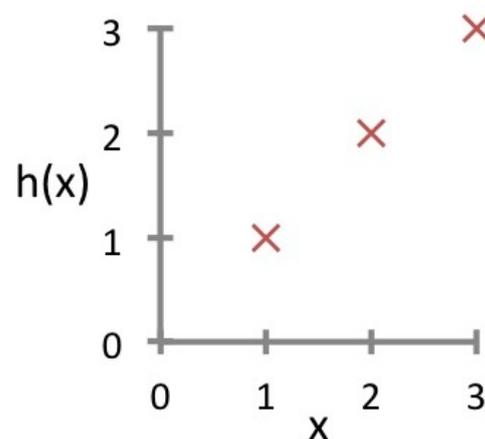


5. Suponha que temos um conjunto de treinamento com  $m = 3$  exemplos, plotados abaixo. Nossa função hipótese é  $h_\theta(x) = \theta_1 x$ , com parâmetro  $\theta_1$ . A função custo  $J(\theta_1)$  é dada por:

$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

Qual o valor de  $J(0)$ ?

- a. 0
- b.  $\frac{1}{6}$
- c. 1
- d.  $\frac{14}{6}$



## Parte III

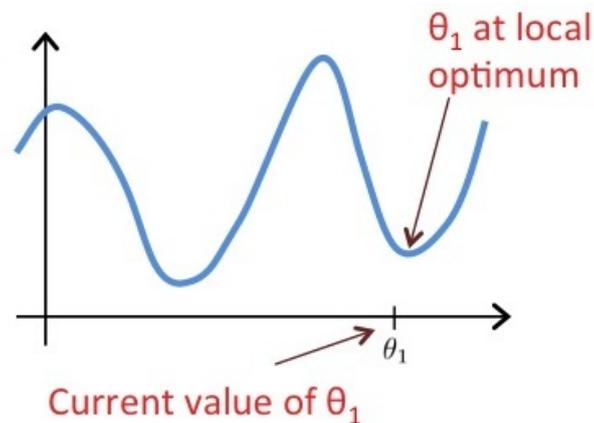
# Parâmetro de aprendizagem

6. Suponha  $\theta_0 = 1$ ,  $\theta_1 = 2$ , e atualizamos automaticamente  $\theta_0$  e  $\theta_1$  usando a regra à seguir:

$$\theta_j := \theta_j + \sqrt{\theta_0 \theta_1} \quad (\text{for } j = 0 \text{ and } j = 1)$$

Quais os valores resultantes de  $\theta_0$  e  $\theta_1$ ?

- a.  $\theta_0 = 1$ ,  $\theta_1 = 2$
  - b.  $\theta_0 = 1 + \sqrt{2}$ ,  $\theta_1 = 2 + \sqrt{2}$
  - c.  $\theta_0 = 2 + \sqrt{2}$ ,  $\theta_1 = 1 + \sqrt{2}$
  - d.  $\theta_0 = 1 + \sqrt{2}$ ,  $\theta_1 = 2 + \sqrt{2 \cdot (1 + \sqrt{2})}$
7. Suponha  $\theta_1$  é um mínimo local de  $J(\theta_1)$ , como é mostrado na figura.

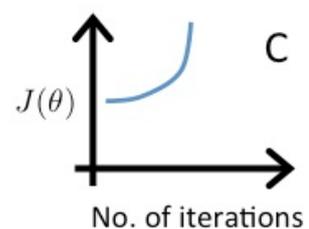
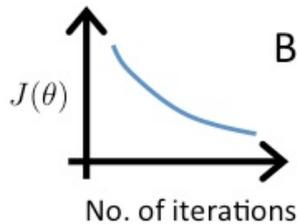
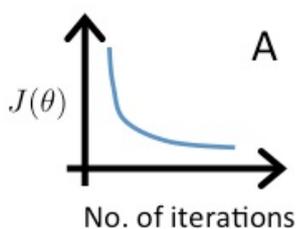


- O que a etapa de descida de gradiente  $\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$  vai fazer?
- a. Mantém  $\theta_1$  inalterado.
  - b. Altera  $\theta_1$  em uma direção aleatória.
  - c. Move  $\theta_1$  na direção do mínimo global de  $J(\theta_1)$
  - d. Diminuí o valor de  $\theta_1$
8. Quais das seguintes afirmações são verdadeiras? Selecione todas que se aplicam.
- a. Para fazer convergir a descida do gradiente, devemos diminuir lentamente  $\alpha$  com o tempo.
  - b. O gradiente descendente é garantido para encontrar o mínimo global para qualquer função  $J(\theta_0, \theta_1)$ .
  - c. A descida do gradiente pode convergir mesmo se  $\alpha$  for mantido fixa. (Mas  $\alpha$  não pode ser muito grande, caso contrário, pode falhar na convergência.)
  - d. Para a escolha específica da função de custo  $J(\theta_0, \theta_1)$  usada na regressão linear, não há ótimo local (além do ótimo global).

## Parte IV

# Regressão multilinear (multivariada linear)

9. Suponha que você esteja usando um algoritmo de aprendizado para estimar o preço das casas em uma cidade. Você deseja que uma de suas características  $x_i$  capture a idade da casa. Em seu conjunto de treinamento, todas as suas casas têm uma idade entre 30 e 50 anos, com uma idade média de 38 anos. Qual das opções a seguir você usaria como recursos, supondo que use dimensionamento de recursos e normalização média?
- a.  $x_i = \text{idade da casa}$
  - b.  $x_i = \frac{\text{idade da casa}}{50}$
  - c.  $x_i = \frac{\text{idade da casa} - 38}{50}$
  - d.  $x_i = \frac{\text{idade da casa} - 38}{20}$
10. Suponha que um amigo fizesse uma descida gradiente três vezes, com  $\alpha = 0.01$ ,  $\alpha = 0.1$ , e  $\alpha = 1$ , e obtivesse os três gráficos a seguir (rotulados A, B, e C):



Quais gráficos correspondem a quais valores de  $\alpha$ ?

- a. A é  $\alpha = 0.01$ , B é  $\alpha = 0.1$  e C é  $\alpha = 1$ .
- b. A é  $\alpha = 0.1$ , B é  $\alpha = 0.01$  e C é  $\alpha = 1$ .
- c. A é  $\alpha = 1$ , B é  $\alpha = 0.01$  e C é  $\alpha = 0.1$ .
- d. A é  $\alpha = 1$ , B é  $\alpha = 0.1$  e C é  $\alpha = 0.01$ .

**Feedback:** No gráfico C, a função de custo está aumentando, portanto, a taxa de aprendizado está definida como muito alta. Ambos os gráficos A e B convergem para um ótimo da função de custo, mas o gráfico B o faz muito lentamente, então sua taxa de aprendizado é definida muito baixa. O gráfico A fica entre os dois.

11. Suponha que você queira prever o preço de uma casa em função de seu tamanho. Seu modelo é

$$h_{\theta} = \theta_0 + \theta_1(\text{tamanho}) + \theta_2\sqrt{\text{tamanho}}$$

Suponha que o tamanho varia de 1 a 1000 (feet<sup>2</sup>). Você vai implementar isso ajustando um modelo

$$h_{\theta}(x) = \theta_0 + \theta_1x_1 + \theta_2x_2$$

Por fim, suponha que você queira usar o dimensionamento de recursos (sem normalização média).

Qual das seguintes opções para  $x_1$  e  $x_2$  você deve usar? (Nota:  $\sqrt{1000} \approx 32$ ).

- a.  $x_1 = \text{tamanho}, x_2 = 32\sqrt{\text{tamanho}}$
- b.  $x_1 = 32(\text{tamanho}), x_2 = \sqrt{\text{tamanho}}$
- c.  $x_1 = \frac{\text{tamanho}}{1000}, x_2 = \frac{\sqrt{\text{tamanho}}}{32}$
- d.  $x_1 = \frac{\text{tamanho}}{32}, x_2 = \sqrt{\text{tamanho}}$

## Parte V

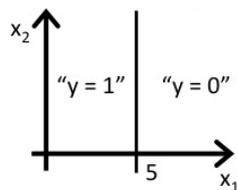
# Classificação e representação

12. Suponha que queiramos prever, a partir dos dados  $x$  sobre um tumor, se ele é maligno ( $y = 1$ ) ou benigno ( $y = 0$ ). Nosso classificador de regressão logística produz, para um tumor específico,  $h_\theta(x) = P(y = 1|x; \theta) = 0.7$ , então estimamos que há uma chance de 70% de este tumor ser maligno. Qual deve ser nossa estimativa para  $P(y = 0|x; \theta)$ , a probabilidade de o tumor ser benigno?

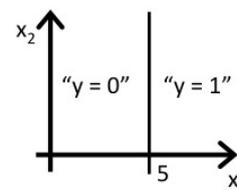
- a.  $P(y = 0|x; \theta) = 0.3$                        c.  $P(y = 0|x; \theta) = 0.7^2$   
 b.  $P(y = 0|x; \theta) = 0.7$                        d.  $P(y = 0|x; \theta) = 0.3 \times 0.7$

13. Considere a regressão logística com dois recursos  $x_1$  e  $x_2$ . Suponha que  $\theta_0 = 5$ ,  $\theta_1 = -1$ ,  $\theta_2 = 0$ , de modo que  $h_\theta(x) = g(5 - x_1)$ . Qual destes mostra a fronteira de decisão de  $h_\theta(x)$ ?

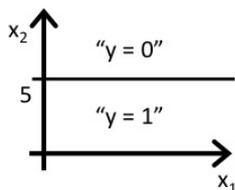
a.



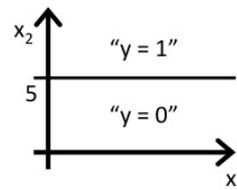
c.



b.



d.



## Parte VI

# Modelo de regressão logística

14. Suponha que você tenha um problema de classificação de várias classes com  $k$  classes (onde  $y \in \{1, 2, \dots, k\}$ ). Usando o método 1-vs.-all, quantos classificadores de regressão logística diferentes você acabará treinando?

- a.  $k - 1$                        c.  $k + 1$   
 b.  $k$                        d. Aproximadamente  $\log_2(k)$

## Parte VII

# Resolvendo o problema do Overfitting

15. Considere o problema do diagnóstico médico de classificar os tumores como malignos ou benignos. Se uma hipótese  $h_\theta(x)$  sobreajustar o conjunto de treinamento, isso significa que:
- a. Faz previsões precisas para exemplos no conjunto de treinamento e generaliza bem para fazer previsões precisas sobre novos exemplos nunca vistos.
  - b. Não faz previsões precisas para exemplos no conjunto de treinamento, mas generaliza bem para fazer previsões precisas sobre novos exemplos nunca vistos anteriormente.
  - c. Faz previsões precisas para exemplos no conjunto de treinamento, mas não generaliza bem para fazer previsões precisas sobre novos exemplos nunca vistos.
  - d. Ele não faz previsões precisas para exemplos no conjunto de treinamento e não generaliza bem para fazer previsões precisas sobre novos exemplos nunca vistos.
16. Na regressão linear regularizada, escolhemos  $\theta$  para minimizar:

$$J(\theta) = \frac{1}{2m} \left[ \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

O que vai acontecer se  $\lambda$  for definido com um valor extremamente grande (talvez grande demais para o nosso problema, digamos  $\lambda = 10^{10}$ )?

- a. O algoritmo funciona bem; definir  $\lambda$  como muito grande não pode prejudicá-lo.
- b. O algoritmo falha em eliminar o overfitting.
- c. O algoritmo resulta em subajuste (falha em ajustar até mesmo o conjunto de treinamento).
- d. A descida gradiente não convergirá.