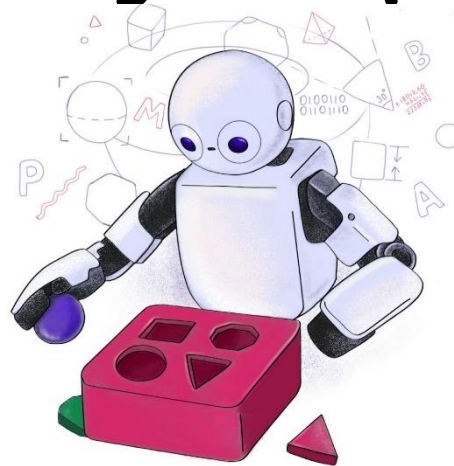


T320 - Introdução ao Aprendizado de Máquina II: *Classificação (Parte IV)*



Inatel

Felipe Augusto Pereira de Figueiredo
felipe.figueiredo@inatel.br

Recapitulando

- Anteriormente, aprendemos uma nova função de limiar, chamada de ***função logística***, com a qual foi possível se encontrar uma solução com ***gradiente descendente***.
- Classificadores que utilizam a ***função logística*** como ***função de limiar*** são conhecidos como ***regressores logísticos*** e são utilizados em problemas de ***classificação binária***, ou seja, problemas com 2 classes apenas.
- Na sequência, veremos como lidar com problemas de classificação que envolvem mais de 2 classes, também chamados de classificação multi-classes.

Casos multi-classe

- Até agora, nós vimos como classificar utilizando **regressão logística** quando os dados pertencem a apenas 2 classes (i.e., $Q = 2$), mas e quando existem mais de 2 classes (i.e., $Q > 2$)? Por exemplo
 - Reconhecimento de dígitos escritos à mão: 10 dígitos.
 - Classificação de texto: Esportes, Economia, Política, Entretenimento, etc.
 - Classificação de sentimentos: Neutro, Positivo, Negativo.
- Existem algumas abordagens para classificação multi-classe:
 - Um-contra-o-Resto
 - Um-contra-Um
 - Regressão softmax
- As 2 primeiras abordagens podem ser aplicadas a qualquer tipo de **classificador binário** e não apenas ao **regressor logístico**.
- A terceira abordagem é uma generalização do classificador logístico para problemas multi-classe.

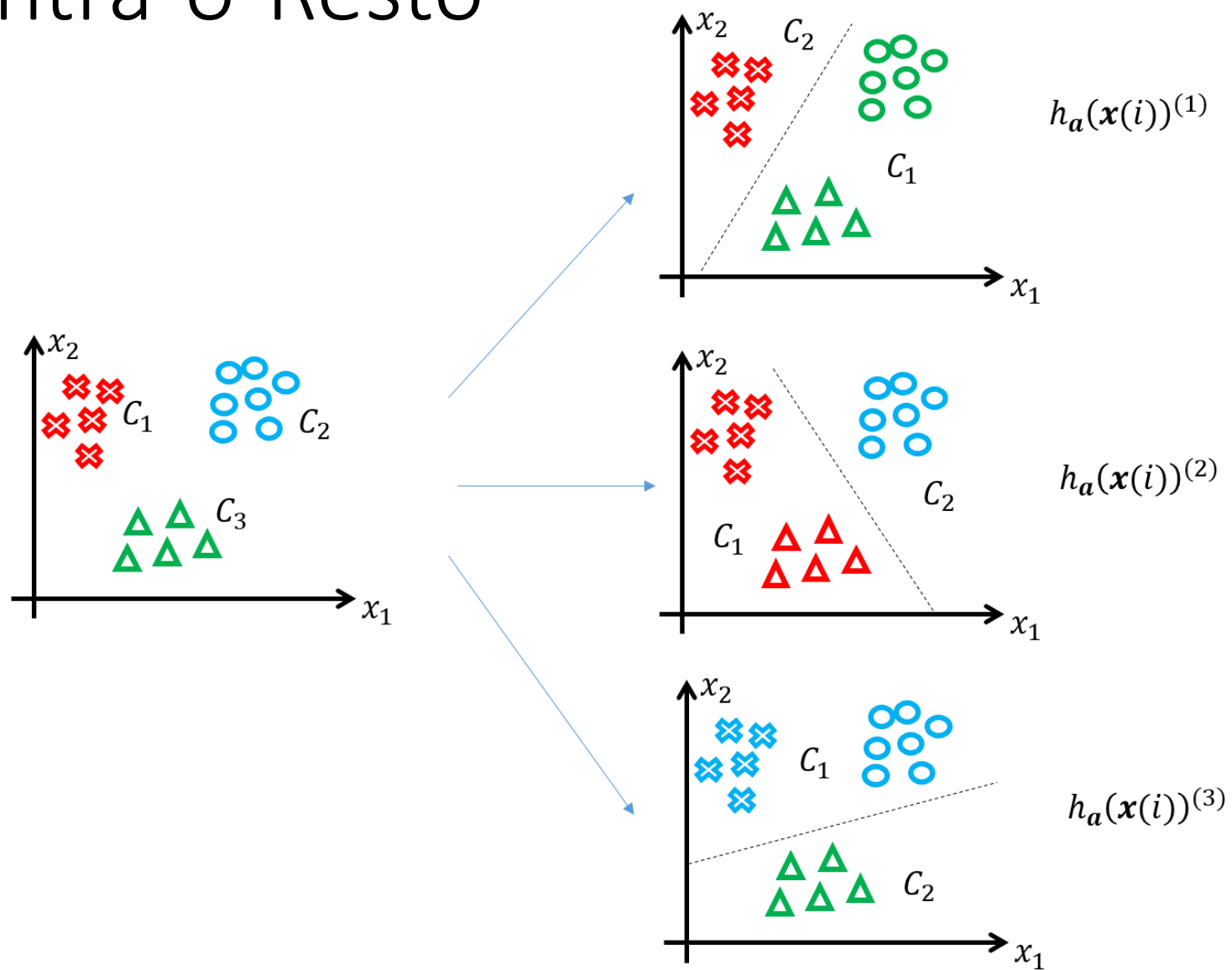
Um-Contra-o-Resto

- Nesta abordagem, nós treinamos um **classificador binário** (e.g., **regressor logístico**), representado por sua função hipótese, $h_a(\mathbf{x}(i))^{(q)}$, para cada classe q para predizer a probabilidade de $\hat{y} = q$, ou seja, $P(\hat{y} = q | \mathbf{x}; \mathbf{a})$.
- Em outras palavras, cria-se Q **classificadores binários**, onde para cada classificador, a classe positiva $C_2 = q$ e a classe negativa C_1 é a junção de todas as outras $Q - 1$ classes.
- Portanto, o **classificador** deve indicar a classe positiva caso o exemplo pertença à classe q , e a classe negativa caso o exemplo pertença a qualquer outra classe.
- Para cada novo exemplo de entrada, \mathbf{x} , realiza-se as predições e escolhe-se a classe que maximize

$$C_q = \arg \max_q h_a(\mathbf{x}(i))^{(q)}.$$

- A vantagem desta abordagem é que se treina apenas Q **classificadores**.
- A desvantagem é que cada **classificador binário** precisa ser treinado com um conjunto negativo que é $Q-1$ vezes maior, o que pode aumentar o tempo de treinamento.

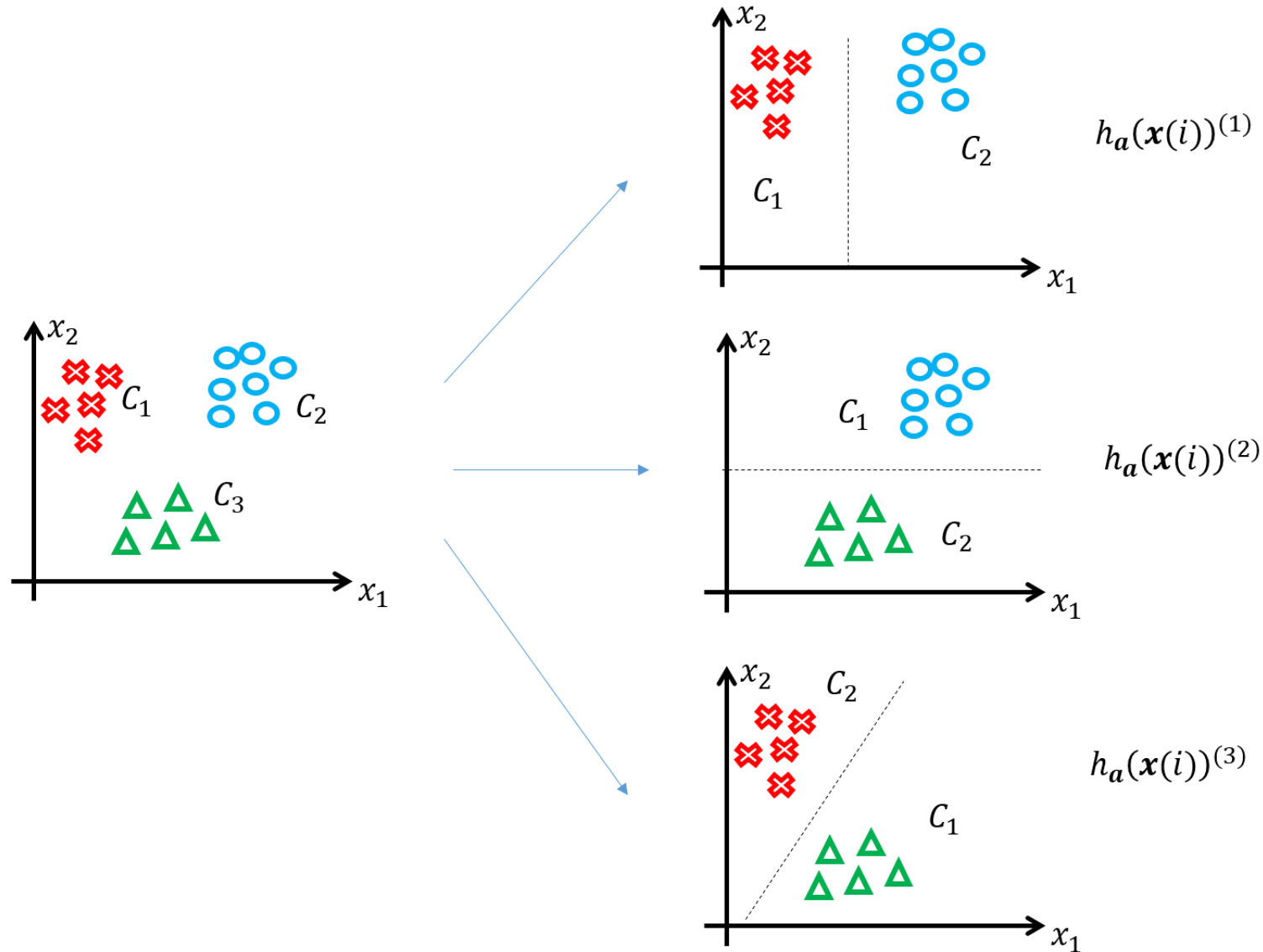
Um-Contra-o-Resto



Um-Contra-Um

- Nesta abordagem, treina-se $Q(Q - 1)/2$ **classificadores binários**.
- Cada **classificador** é construído para fazer a distinção entre exemplos pertencentes a cada um dos possíveis **pares** de classes.
 - Se $Q = 4$, então treina-se 6 **classificadores** para classificar entre C_1/C_2 , C_1/C_3 , C_1/C_4 , C_2/C_3 , C_2/C_4 , e C_3/C_4 .
- No final, cada exemplo é classificado conforme o **voto majoritário** entre os **classificadores**.
- A principal vantagem da abordagem **Um-Contra-Um** é que cada **classificador** precisa ser treinado apenas na parte do conjunto de treinamento para as duas classes que ele deve distinguir.
- A desvantagem é que por exemplo, se $Q = 10$, temos que treinar 45 **classificadores**.

Um-Contra-Um



Regressão Softmax

- Também conhecida como ***regressão logística multinomial***.
- A ideia é ter um ***único*** classificador que classifique mais de 2 classes.
 - Por exemplo, para um problema com 4 classes, teríamos um único classificador, mas com 4 saídas.
- É importante salientar que ele prediz ***apenas uma classe de cada vez***, ou seja, ele é ***multi-classe*** e não ***multi-saída***, portanto, ele deve ser usado apenas com ***classes mutuamente exclusivas***, como por exemplo diferentes tipos de plantas, dígitos, categorias de notícias, etc.
- Portanto, você não poderia usá-lo para reconhecer várias pessoas em uma foto, por exemplo.
- É uma abordagem mais robusta que as anteriores e que consiste em criar um modelo em que cada saída representa a ***probabilidade*** de um exemplo pertencer a uma classe específica.

Regressão Softmax

- Isto é feito a partir de uma generalização da **regressão logística** chamada de **função softmax**, a qual é definida como

$$P(C_q | \mathbf{x}(i)) = h_a^q(\mathbf{x}(i)) = \frac{e^{g_q(\mathbf{x}(i))}}{\sum_{j=1}^Q e^{g_j(\mathbf{x}(i))}} = \frac{e^{\mathbf{x}(i)^T \mathbf{a}_q}}{\sum_{j=1}^Q e^{\mathbf{x}(i)^T \mathbf{a}_j}} \in \mathbb{R} [0,1],$$

O somatório de termos exponenciais normaliza o valor da q -ésima saída de tal forma que o somatório das Q saídas seja igual a 1.

onde $\mathbf{a}_q = [a_0^q, a_1^q, \dots, a_K^q]^T$ é o **vetor de pesos** associado à q -ésima saída do classificador, $h_a^q(\mathbf{x}(i))$ é a **função hipótese** associada à q -ésima classe e

$$g_q(\mathbf{x}(i)) = \mathbf{x}(i)^T \mathbf{a}_q = a_0^q + a_1^q x_1 + \dots + a_K^q x_K = a_0^q + \sum_{k=1}^K a_k^q x_k,$$

é a **função discriminante** para a q -ésima classe.

- A **função softmax** estende a ideia do **regressor logístico** ao mundo multi-classes.
- Ou seja, a função softmax atribui probabilidades, no intervalo $[0, 1]$, a cada classe em um problema com várias classes.
- Essas probabilidades devem somar 1.

Regressão Softmax

- Assim como fizemos anteriormente, precisamos definir uma **função de erro** para podermos encontrar os **pesos** das **Q funções hipótese** do classificador.
- A **função de erro médio** é dada por

$$J_e(A) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{q=1}^Q 1\{y(i) + 1 == q\} \log(h_a^q(x(i))),$$

onde $1\{\cdot\}$ é a **função indicadora**, de modo que $1\{\text{uma condição verdadeira}\} = 1$ e $1\{\text{uma condição falsa}\} = 0$ e $A \in \mathbb{R}^{K+1 \times Q}$ é a matriz com os **pesos** para todas as **funções hipótese** das Q classes.

- Usando-se a representação **one-hot-encoding**, a equação acima pode ser re-escrita como

$$J_e(A) = -\frac{1}{N} \sum_{i=0}^{N-1} \mathbf{y}(i)^T \log(\mathbf{h}_a(x(i))),$$

onde $\mathbf{y}(i) = [1\{y(i) + 1 == 1\}, \dots, 1\{y(i) + 1 == Q\}]^T$ é o vetor com **one-hot-encoding** e

$$\mathbf{h}_a(x(i)) = [h_a^1(x(i)), \dots, h_a^Q(x(i))]^T = [P(C_1 | \mathbf{x}; \mathbf{a}_1) \dots P(C_Q | \mathbf{x}; \mathbf{a}_Q)]^T.$$

- Observem que, quando existem apenas duas classes ($Q = 2$), a **função de erro** acima é equivalente à **função de erro** do **regressor logístico**.

Regressão Softmax

- Usamos o algoritmo do **gradiente descendente** para encontrar os **pesos** que **minimizam** a **função de erro médio**.

- A atualização iterativa dos **pesos** ae q -ésima classe é dada por

$$\mathbf{a}_q = \mathbf{a}_q - \alpha \frac{\partial J_e(A)}{\partial \mathbf{a}_q}$$

- A derivada de $J_e(A)$ com respeito a cada vetor de pesos, \mathbf{a}_q , tem uma expressão semelhante àquela obtida para a **regressão logística**:

$$\frac{\partial J_e(A)}{\partial \mathbf{a}_q} = -\frac{1}{N} \sum_{i=0}^{N-1} [y^q(i) - h_a^q(\mathbf{x}(i))] \mathbf{x}(i)^T.$$

- $\sum_{q=1}^Q h_a^q(\mathbf{x}(i)) = \sum_{q=1}^Q P(C_q | \mathbf{x}; \mathbf{a}_q) = 1$, ou seja, o somatório da **probabilidade condicional** de todas as classes é igual a 1.

- $0 \leq h_a^q(\mathbf{x}(i)) \leq 1$, ou seja, temos, um vetor

$$\mathbf{h}_a(\mathbf{x}(i)) = [h_a^1(\mathbf{x}(i)) \quad \dots \quad h_a^Q(\mathbf{x}(i))] \in \mathbb{R}^{Q \times 1}$$

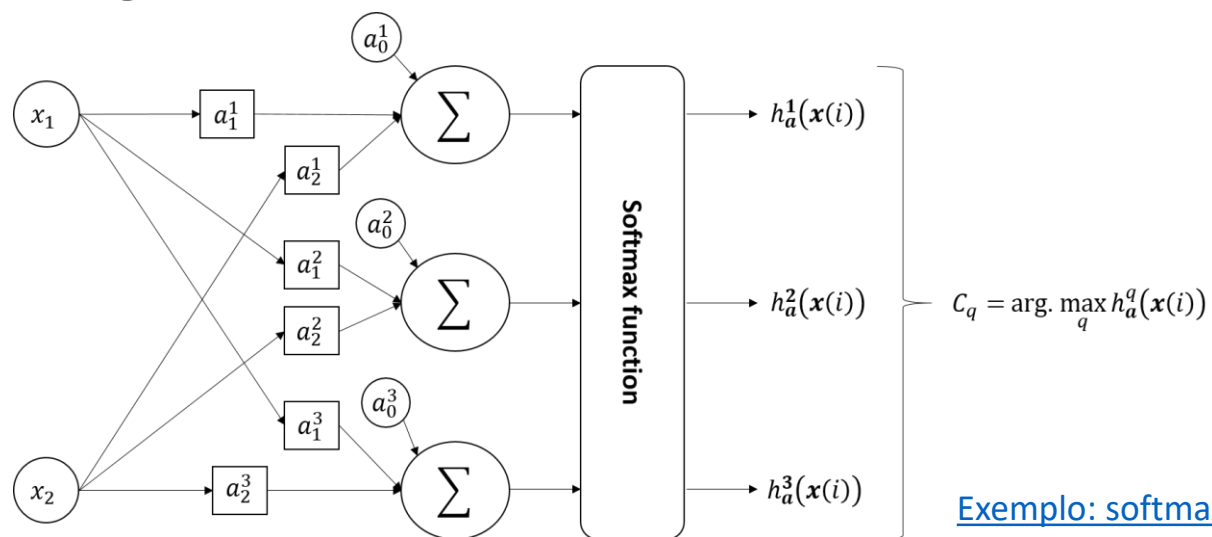
que atende os requisitos de uma **função probabilidade de massa** (PMF, do inglês **probability mass function**) **multinomial**.

Regressão Softmax

- Após o treinamento do classificador, para cada novo exemplo de entrada, \mathbf{x} , realiza-se as previsões e escolhe-se a classe que maximize

$$C_q = \arg. \max_q h_a^q(\mathbf{x}(i)) = \arg. \max_q P(C_q \mid \mathbf{x}(i); \mathbf{a}_q) = \arg. \max_q \mathbf{x}(i)^T \mathbf{a}_q .$$

- Assim como o classificador de regressão logística, o classificador de regressão softmax prevê a classe com a maior probabilidade estimada (que é simplesmente a classe com a maior valor para o produto escalar $\mathbf{x}(i)^T \mathbf{a}_q$).
- A arquitetura de um regressor softmax é mostrada abaixo.



Tarefas

- **Quiz:** “*T320 - Quiz - Classificação (Parte IV)*” que se encontra no MS Teams.
- **Exercício Prático:** [Laboratório #4](#).
 - Pode ser baixado do MS Teams ou do GitHub.
 - Pode ser respondido através do link acima (na nuvem) ou localmente.
 - [Instruções para resolução e entrega dos laboratórios](#).
 - **Atividades podem ser feitas em grupo, mas as entregas devem ser individuais.**

Obrigado!