

EA991 - Laboratório de Aprendizado de Máquina





Classificação

Prof. Denis G. Fantinato
Prof. Levy Boccato



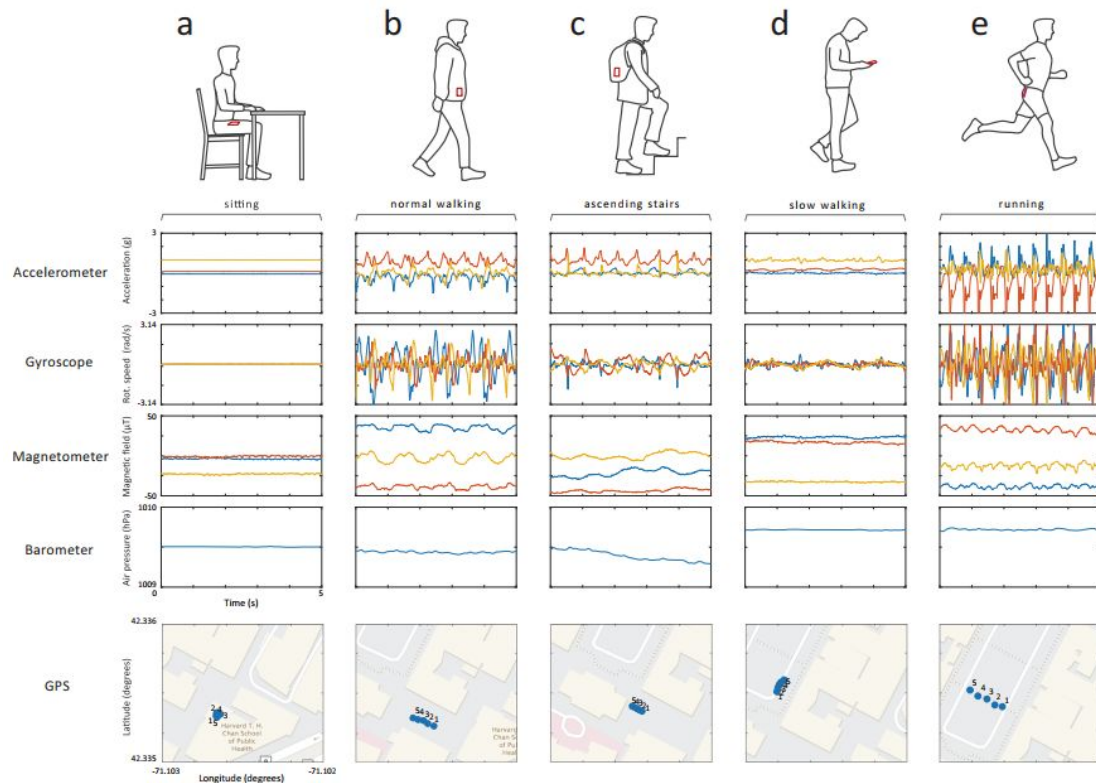
Introdução

- **Classificação** é a tarefa de atribuir a cada amostra ou exemplo de entrada um rótulo correspondente à classe à qual aquele dado pertence.
- **Exemplo:** detecção de *posts* relevantes em redes sociais

Event	Positive Images	Negative Images
National Museum		
Notre-Dame Cathedral		

Introdução

- Exemplo: reconhecimento de atividades humanas



Cenários de classificação

Classificação binária



- Spam
- Not spam

Classificação multi-classe



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

Classificação multirrótulo



- Dog
- Cat
- Horse
- Fish
- Bird
- ...

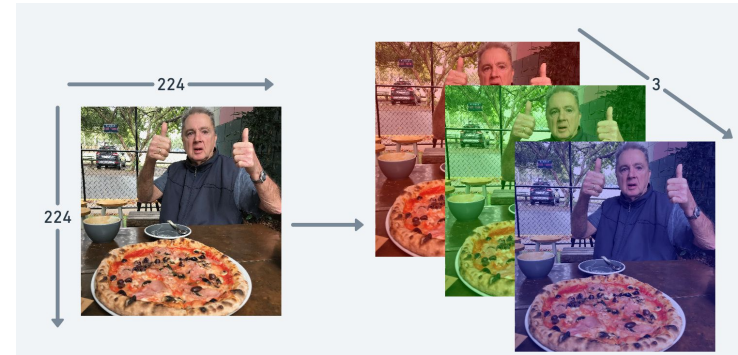
Classificação

- Conjunto de dados (*dataset*): $\{\mathbf{x}(i); \mathbf{y}(i)\}_{i=0}^{N-1}$
 - Cada amostra é caracterizada por K atributos e vem acompanhada da saída esperada.
 - Existem ao todo N amostras rotuladas.

- **Entrada:**

Renda mensal	Idade	Escolaridade	...	Altura
1.800,00	19	Médio	⋮	1,78
10.500,00	32	Superior	⋮	1,69
6.870,00	43	Superior	...	1,99

Vetor ou lista de K elementos



Tensor: (3,224,224)

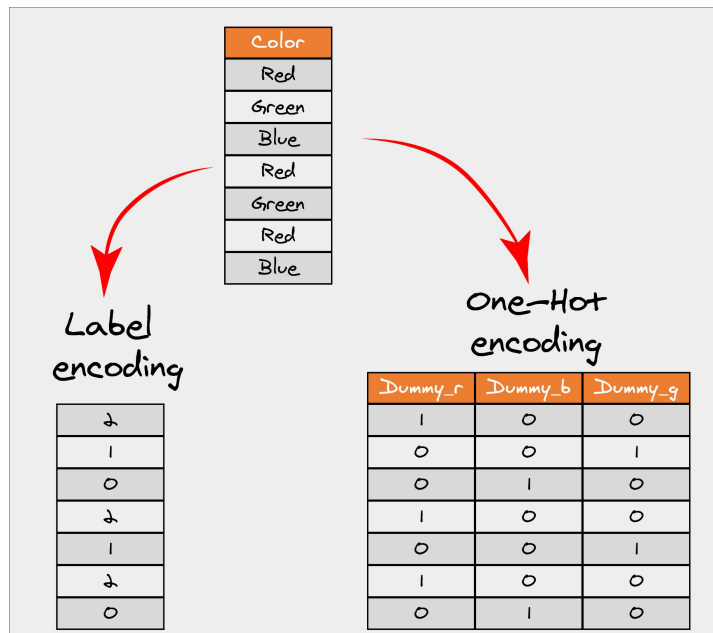
Classificação

- Saída desejada:

- Caso binário:

$$y(i) = \begin{cases} 0, & \mathbf{x}(i) \in C_1 \\ 1, & \mathbf{x}(i) \in C_2 \end{cases}$$

- Multi-classe: neste caso, o classificador produz múltiplas saídas, cada uma representando a possibilidade de o padrão pertencer a uma classe específica.





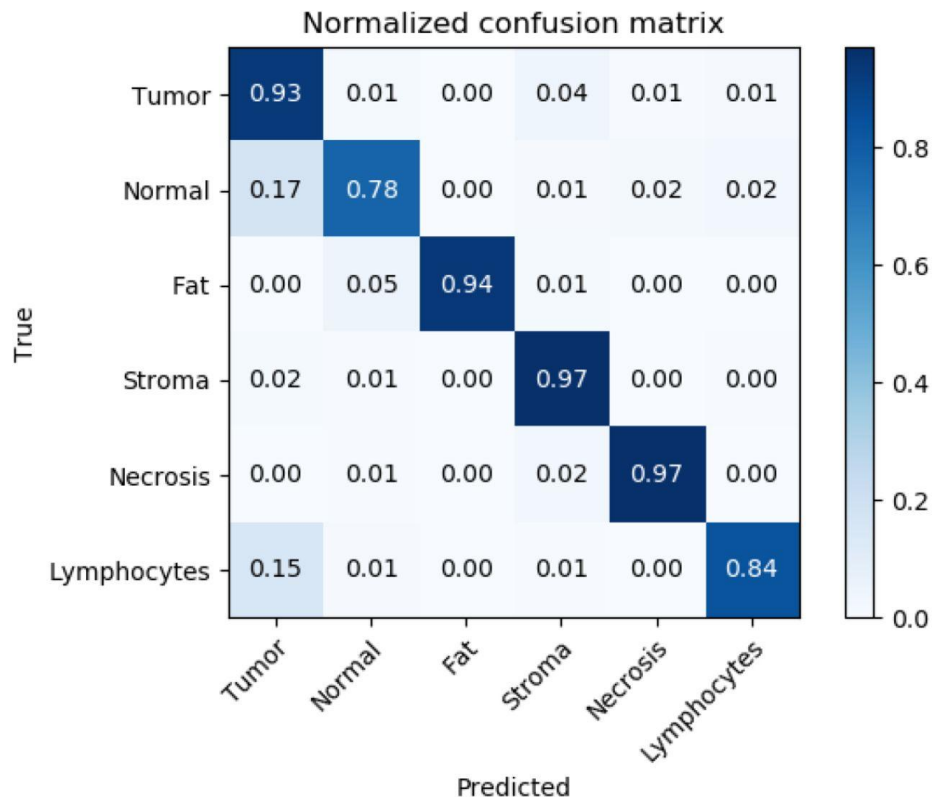
Métricas de avaliação

- **Matriz de confusão:** contabiliza o número de classificações corretas e incorretas para cada uma das Q classes existentes. O elemento c_{ij} indica quantos padrões da classe i foram designados à classe j . Em sua diagonal, portanto, encontramos o número de classificações corretas.
 - Cada linha está associada a uma classe verdadeira;
 - Cada coluna está associada a uma classe estimada.

		Classe estimada	
		+	-
Classe verdadeira	+	Verdadeiro positivo (TP)	Falso negativo (FN)
	-	Falso positivo (FP)	Verdadeiro negativo (TN)

Métricas de avaliação

- Matriz de confusão:





Métricas de avaliação

- **Taxa de falso positivo:** proporção de exemplos da classe negativa (-) classificados incorretamente.

$$\text{Taxa de falso positivo} = \frac{FP}{TN + FP} = \frac{FP}{N_-}$$

- **Taxa de falso negativo:** proporção de exemplos da classe positiva (+) classificados incorretamente.

$$\text{Taxa de falso negativo} = \frac{FN}{TP + FN} = \frac{FN}{N_+}$$



Métricas de avaliação

- **Precisão:** corresponde à proporção de amostras da classe positiva corretamente classificadas em relação a todos os exemplos atribuídos à classe positiva.

$$\text{Precisão}(\hat{y}(\mathbf{x})) = \frac{TP}{TP + FP}$$

- **Sensibilidade (*recall*):** também conhecida como taxa de verdadeiro positivo, a sensibilidade corresponde à proporção de amostras da classe positiva corretamente classificadas.

$$\text{recall}(\hat{y}(\mathbf{x})) = \frac{TP}{TP + FN}$$

- **Especificidade:** também conhecida como taxa de verdadeiros negativos, a especificidade é dada pela proporção de amostras da classe negativa corretamente classificadas.

$$\text{Especificidade}(\hat{y}(\mathbf{x})) = \frac{TN}{TN + FP}$$

Métricas de avaliação

		Classe estimada		
		+	-	
Classe verdadeira	+	Verdadeiro positivo (TP)	Falso negativo (FN)	Recall: $TP / (TP + FN)$
	-	Falso positivo (FP)	Verdadeiro negativo (TN)	Especificidade: $TN / (TN + FP)$
		Precisão (+): $TP / (TP + FP)$	Precisão (-): $TN / (TN + FN)$	Acurácia: $(TP + TN) / (TP + TN + FP + FN)$



Métricas de avaliação

- **F_1 -score:** uma vez que precisão e *recall* costumam ser analisados juntos, existe uma métrica única, denominada F-medida (ou F-score), que combina as duas informações através de uma média harmônica ponderada:

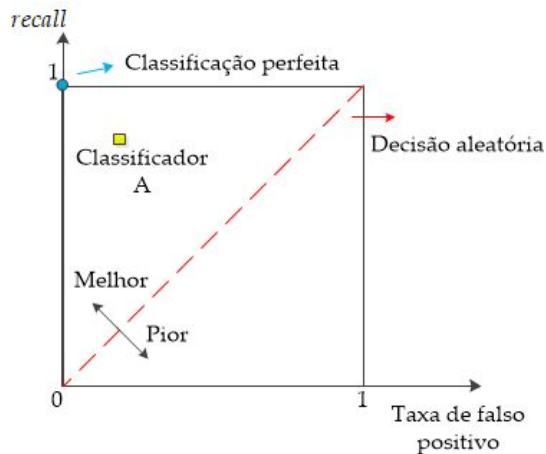
$$F_1 = 2 \frac{\text{recall}(\hat{y}(\mathbf{x})) \times \text{precisão}(\hat{y}(\mathbf{x}))}{\text{recall}(\hat{y}(\mathbf{x})) + \text{precisão}(\hat{y}(\mathbf{x}))}$$

- **Acurácia balanceada:** trata-se de uma métrica competente para medir o desempenho de um classificador em cenários com um significativo desbalanceamento entre as classes, isto é, quando as classes têm quantidades de amostras bem discrepantes. Nela, todas as classes têm a mesma importância, independentemente da quantidade de amostras.

$$\text{BA} = \frac{\text{recall}_1 + \dots + \text{recall}_Q}{Q}$$

Métricas de avaliação

- **Obs.:** É possível estender de maneira natural estas métricas para o cenário multi-classe; para isto, basta tomar, uma vez, cada classe C_k , $k = 1, \dots, Q$, como sendo a classe positiva, enquanto todas as demais classes formam a classe negativa; assim, obtemos os valores das métricas para cada classe.
- **Curva ROC:** trata-se de um gráfico em que a taxa de verdadeiro positivo, a qual equivale ao *recall*, é exibida em função da taxa de falso positivo conforme se altera o limiar (*threshold*) de decisão.



Métodos de classificação





Regressão logística

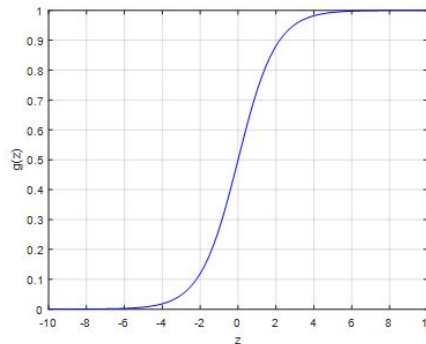
- Trata-se de uma abordagem de classificação que tenta promover a separação das classes com base em fronteiras de decisão lineares.
 - Originalmente é formulada para o caso de classificação binária, mas pode ser estendida para multi-classe.
- **Caso binário:** o modelo produz uma única saída por meio do seguinte mapeamento:

$$\hat{y}(\mathbf{x}) = \frac{e^{(w_0 + w_1 x_1 + \dots + w_K x_K)}}{1 + e^{(w_0 + w_1 x_1 + \dots + w_K x_K)}} = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + \dots + w_K x_K)}}$$

Os coeficientes da combinação linear dos atributos de entrada são os parâmetros ajustáveis do modelo

Regressão logística

- Função logística: $g(z) = \frac{1}{1 + e^{-z}}$



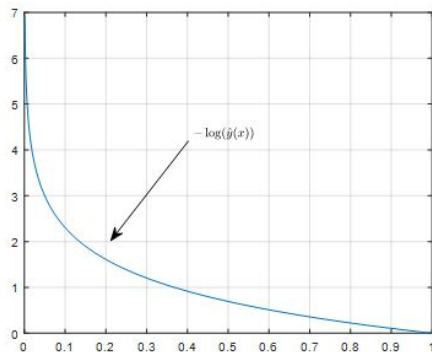
- A regressão logística aplica a função logística ao resultado da combinação linear dos atributos de entrada (mais um termo adicional). Como a saída gerada está sempre entre 0 e 1, ela é interpretada como a *probabilidade de a entrada pertencer à classe positiva*, para a qual a saída desejada é $y = 1$.
- A fronteira de decisão se manifesta quando há uma indeterminação, a saber, quando as probabilidades correspondentes às duas classes são iguais, isto é, 0.5.

$$w_0 + w_1x_1 + \dots + w_Kx_K = 0 \longrightarrow \text{Hiperplano}$$

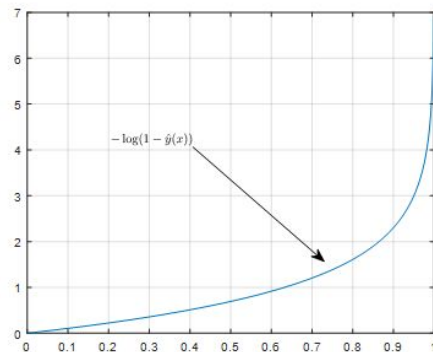
Regressão logística

- Função de perda:

Para uma amostra \rightarrow
$$\text{Custo}(\hat{y}(\mathbf{x}); y) = \begin{cases} -\log \hat{y}(\mathbf{x}), & \text{se } y = 1 \\ -\log(1 - \hat{y}(\mathbf{x})), & \text{se } y = 0 \end{cases}$$



$y = 1$



$y = 0$



Regressão logística

- Função de perda:

Para uma amostra: $\text{Custo}(\hat{y}(\mathbf{x}); y) = \underbrace{-y \log(\hat{y}(\mathbf{x}))}_{\text{Só exerce influência se } y=1} \underbrace{-(1-y) \log(1 - \hat{y}(\mathbf{x}))}_{\text{Penaliza apenas se } y=0}$

Entropia cruzada:

$$J_{\text{CE}}(\mathbf{w}) = -\frac{1}{N} \sum_{i=0}^{N-1} y(i) \log(\hat{y}(\mathbf{x}(i))) + (1 - y(i)) \log(1 - \hat{y}(\mathbf{x}(i)))$$



Regressão logística

- **Treinamento:** é feito com o auxílio de algoritmos iterativos que atualizam os parâmetros \mathbf{w} à medida que os dados são apresentados ao modelo.
- **Ideia base: gradiente descendente**

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \alpha \nabla J_e(\mathbf{w}_i)^T$$

- Cada iteração do algoritmo corresponde a uma atualização do vetor de parâmetros; uma *época* corresponde a uma apresentação completa do conjunto de amostras de treinamento.



Regressão logística

- **Caso multi-classe:**

- A estratégia consiste em montar um modelo que produza Q saídas, tal que cada saída represente a probabilidade de cada padrão pertencer a uma classe específica. Isto pode ser feito a partir da função *softmax*.

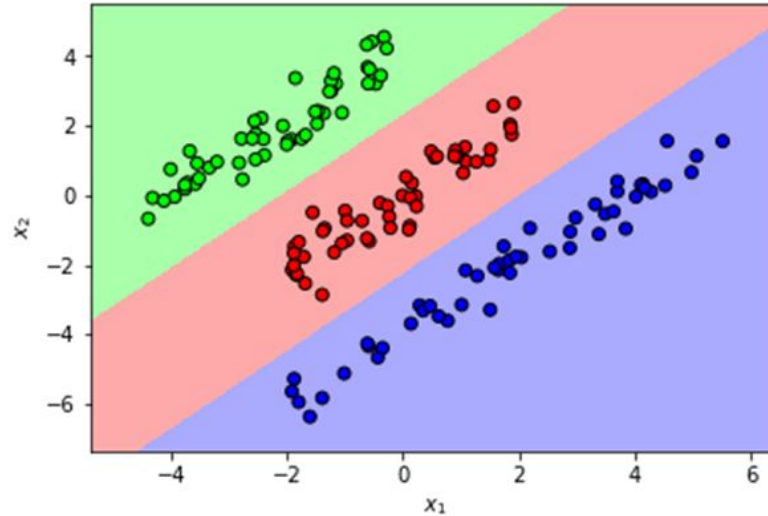
$$P(C_k|\mathbf{x}(i)) = \hat{y}_k(\mathbf{x}(i)) = \frac{e^{(\boldsymbol{\phi}(\mathbf{x}(i))^T \mathbf{w}_k)}}{\sum_j e^{(\boldsymbol{\phi}(\mathbf{x}(i))^T \mathbf{w}_j)}}$$

- Propriedades:

$$\left. \begin{array}{l} \sum_{k=1}^Q \hat{y}_k(\mathbf{x}(i)) = 1 \\ 0 \leq \hat{y}_k(\mathbf{x}(i)) \leq 1 \end{array} \right\} \text{Preenche os requisitos de uma função probabilidade de massa}$$

Regressão logística

- Exemplo:



Fronteiras de decisão obtidas com a regressão logística para um problema com três classes linearmente separáveis



Próximos passos

- Complementar com:
 - scikit-learn e logistic regression (hiperparâmetros)
 - kNN
 - Visão geral e hiperparâmetros
 - SVM
 - Árvore de decisão (overview simples) e *random forest*