

# Aula 1: Introdução a ML e VC

**Lucas Pereira, Rafael Teixeira, Lucas Assis, Anderson Soares**  
Instituto de Informática  
Universidade Federal de Goiás (UFG)

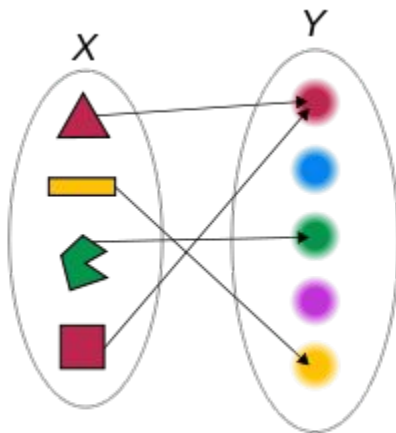
# Sumário

- Introdução ao problema de otimização
- Introdução ao problema de classificação de imagens
- Visão geral de problemas orientados a dados
- Limitações de modelos lineares
- No próximo episódio...

# Introdução ao problema de otimização

- Definição de função

$$f : X \rightarrow Y, \mathbf{y} = f(\mathbf{x})$$

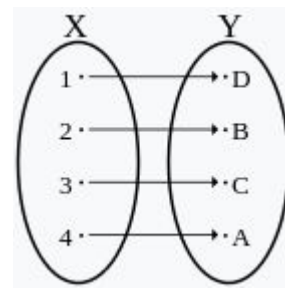
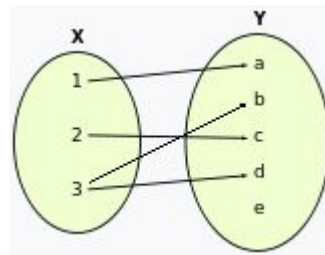
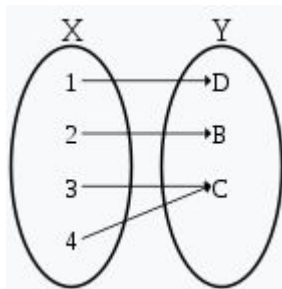
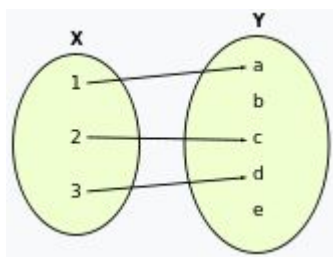
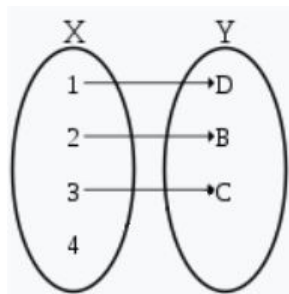


# Introdução ao problema de otimização

- Definição de função

$$f : X \rightarrow Y, \mathbf{y} = f(\mathbf{x})$$

- É ou não é função? Por que?

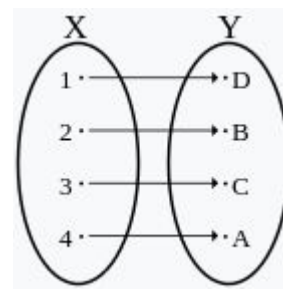
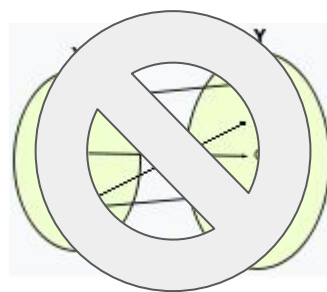
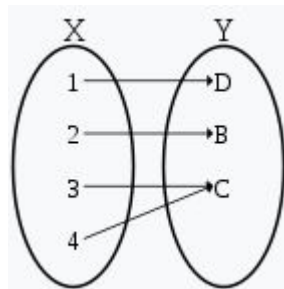
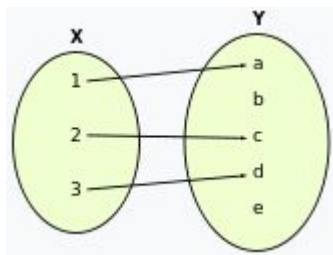
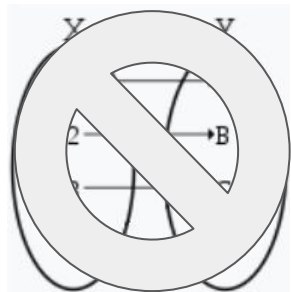


# Introdução ao problema de otimização

- Definição de função

$$f : X \rightarrow Y, \mathbf{y} = f(\mathbf{x})$$

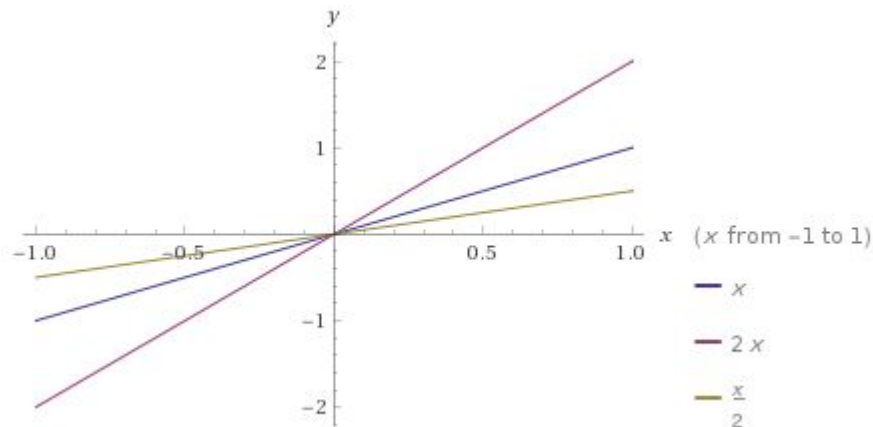
- É ou não é função? Por que?



# Introdução ao problema de otimização

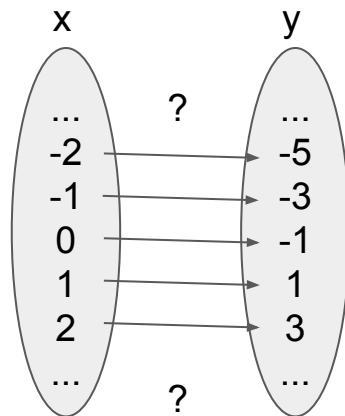
- Definição de função parametrizável

$$f : X, \Theta \rightarrow Y, \mathbf{y} = f(\mathbf{x}; \theta)$$



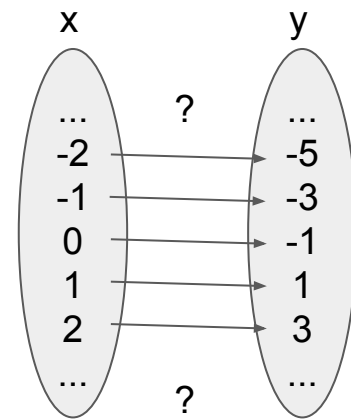
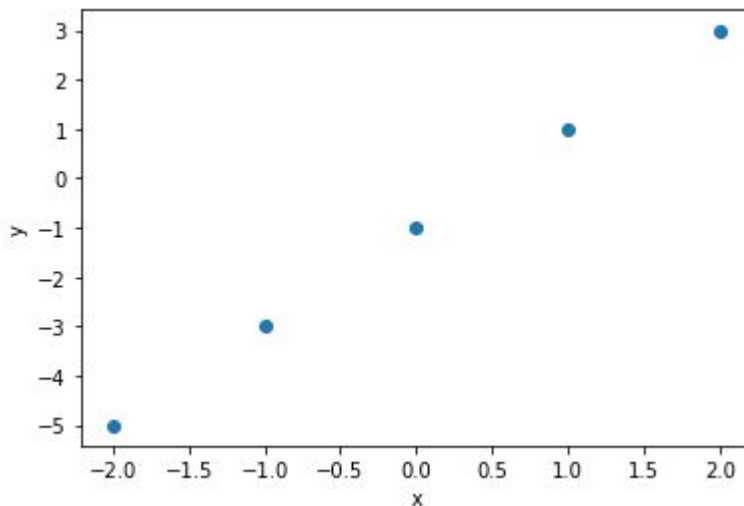
# Introdução ao problema de otimização

- Exemplo de dataset



# Introdução ao problema de otimização

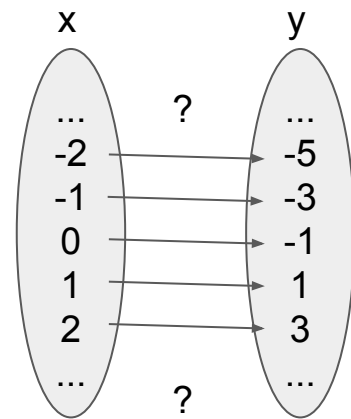
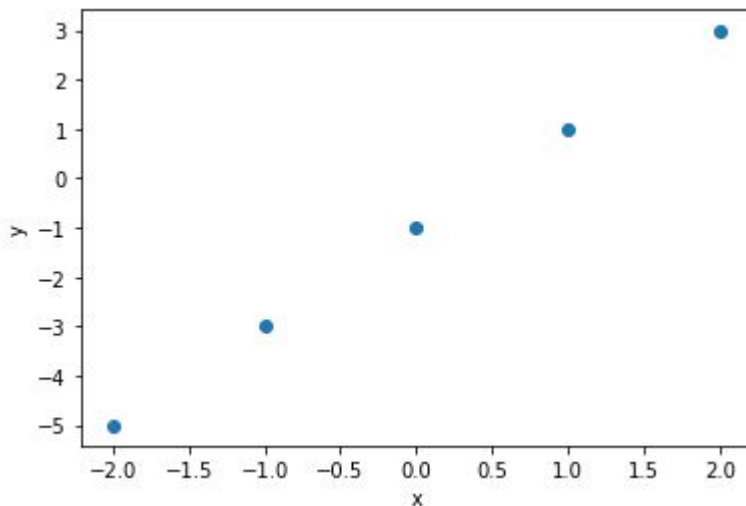
- Visualização dos dados





# Introdução ao problema de otimização

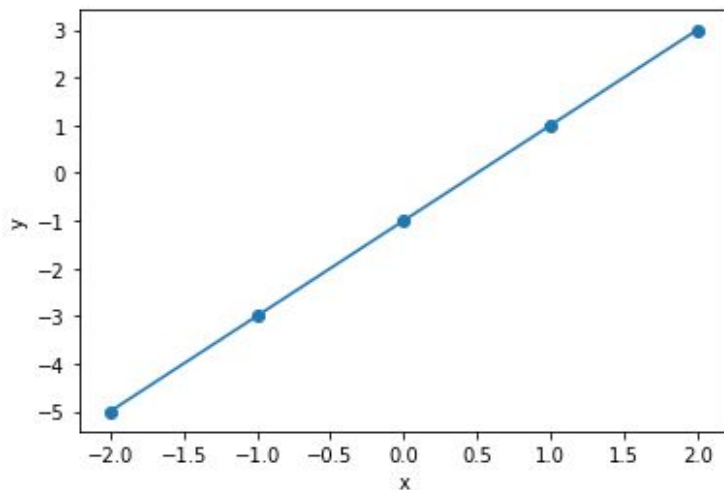
- Regressão linear



$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x; \theta) = \theta_0 x + \theta_1$$

# Introdução ao problema de otimização

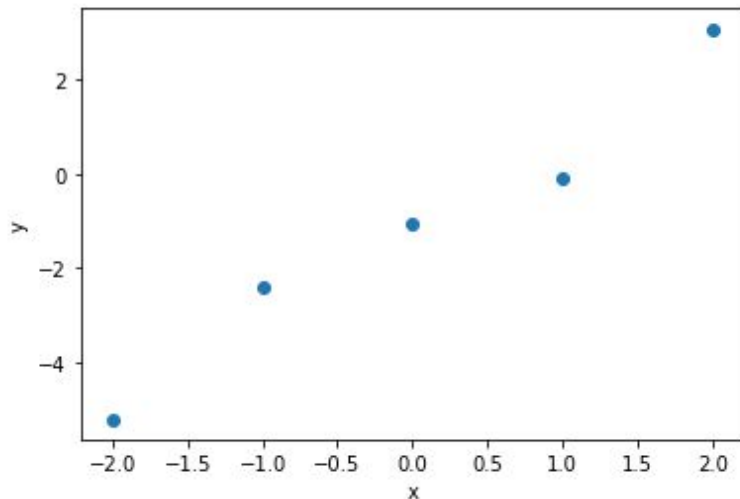
- Regressão linear



$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x; \theta) = 2x - 1$$

# Introdução ao problema de otimização

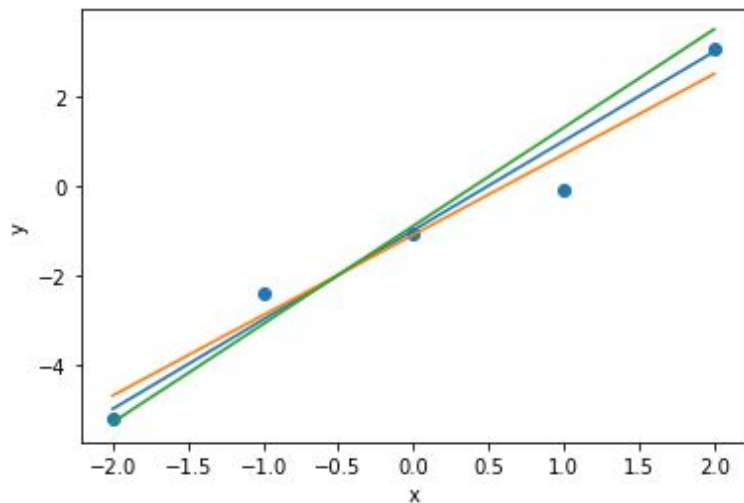
- Regressão linear?



$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x; \theta) = \theta_0 x + \theta_1$$

# Introdução ao problema de otimização

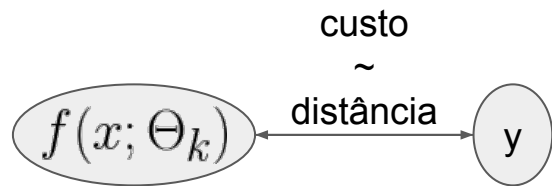
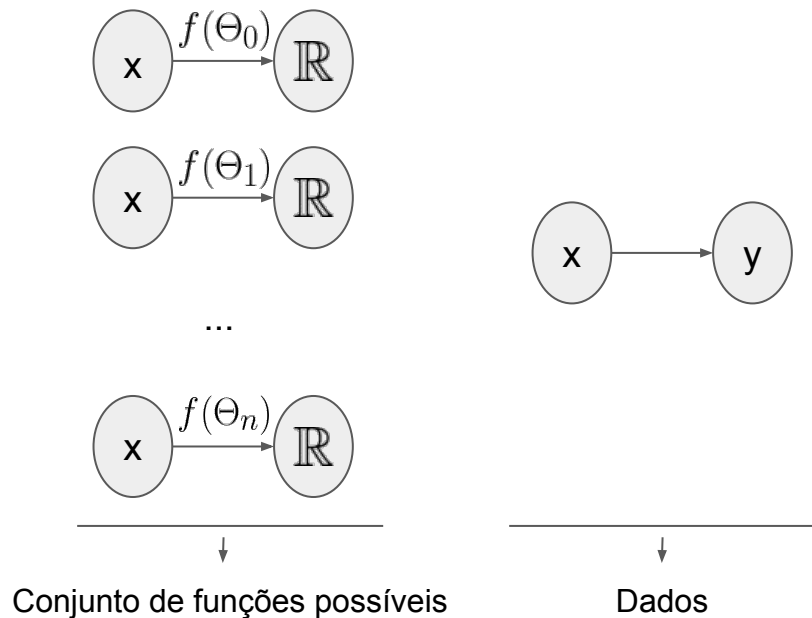
- Regressão linear?



$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x; \theta) = \theta_0 x + \theta_1$$

# Introdução ao problema de otimização

- Função custo (loss)



# Introdução ao problema de otimização

- Distância

$$d : \mathbf{X} \times \mathbf{X} \rightarrow [0, \infty)$$

$$1 : d(x_i, x_j) \geq 0$$

$$2 : d(x_i, x_j) = d(x_j, x_i)$$

$$3 : d(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$$

$$4 : d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$$

# Introdução ao problema de otimização

- Exemplos de distâncias

$$d : \mathbf{X} \times \mathbf{X} \rightarrow [0, \infty)$$

$$d_p(x_i, x_j) = \|x_i - x_j\|_p = \left( \sum_n |x_{i_n} - x_{j_n}|^p \right)^{\frac{1}{p}}$$

$$d_1(x_i, x_j) = \|x_i - x_j\|_1 = \left( \sum_n |x_{i_n} - x_{j_n}| \right)$$

$$d_2(x_i, x_j) = \|x_i - x_j\|_2 = \left( \sum_n |x_{i_n} - x_{j_n}|^2 \right)^{\frac{1}{2}}$$

⋮

$$d_\infty(x_i, x_j) = \|x_i - x_j\|_\infty = \max_n |x_{i_n} - x_{j_n}|$$

# Introdução ao problema de otimização

- Custo

$$d : \mathbf{X} \times \mathbf{X} \rightarrow [0, \infty)$$

$$1 : d(x_i, x_j) \geq a, a \geq 0$$

$$\text{---} 2 : d(x_i, x_j) = d(x_j, x_i) \text{---}$$

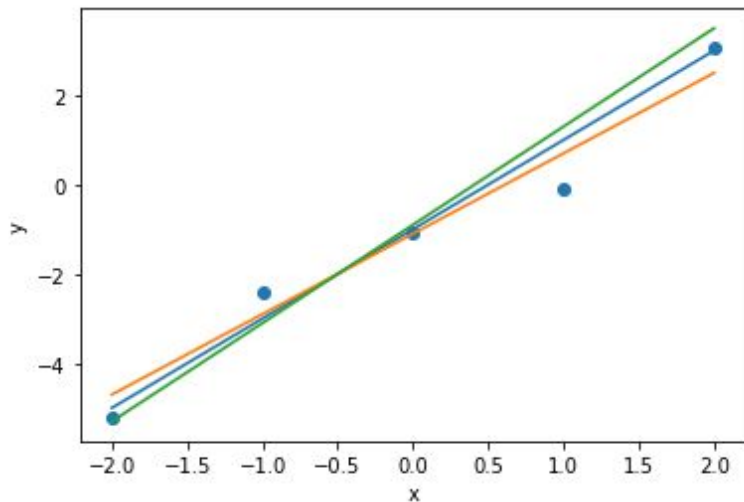
$$3 : d(x_i, x_j) = a \Leftrightarrow x_i = x_j$$

$$\text{---} 4 : d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j) \text{---}$$



# Introdução ao problema de otimização

- Regressão linear!



$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x; \theta) = \theta_0 x + \theta_1$$

$$d : A \times B \rightarrow [0, \infty), d(a, b) = \|a - b\|_2^2$$

$$\mathcal{L}(\Theta) = d(f(x; \theta), \hat{y}) = \|f(x, \theta) - \hat{y}\|_2^2$$

$$\mathcal{L}(\Theta) = \|\theta_0 x + \theta_1 - \hat{y}\|_2^2$$

$$\mathcal{L}(\Theta) = (\theta_0 x + \theta_1 - \hat{y})^2$$

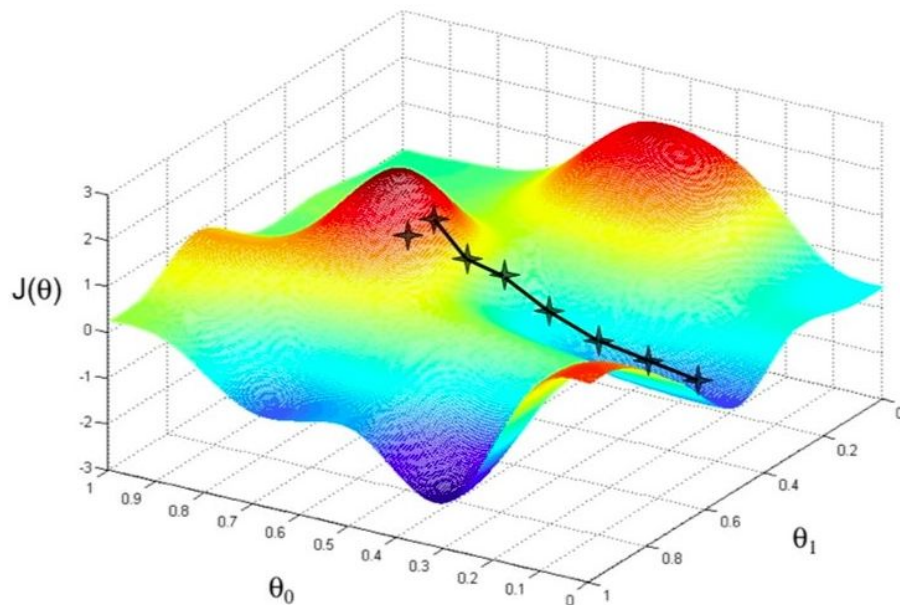
$$\theta = \operatorname{argmin}(\mathcal{L}(\Theta))$$

# Introdução ao problema de otimização

- Otimização: como achar o  $\theta$ ?
  - Solução analítica do problema
  - Busca exaustiva
  - Busca aleatória (ex: algoritmos genéticos)
  - Busca baseada em gradientes (ex: gradiente descendente)

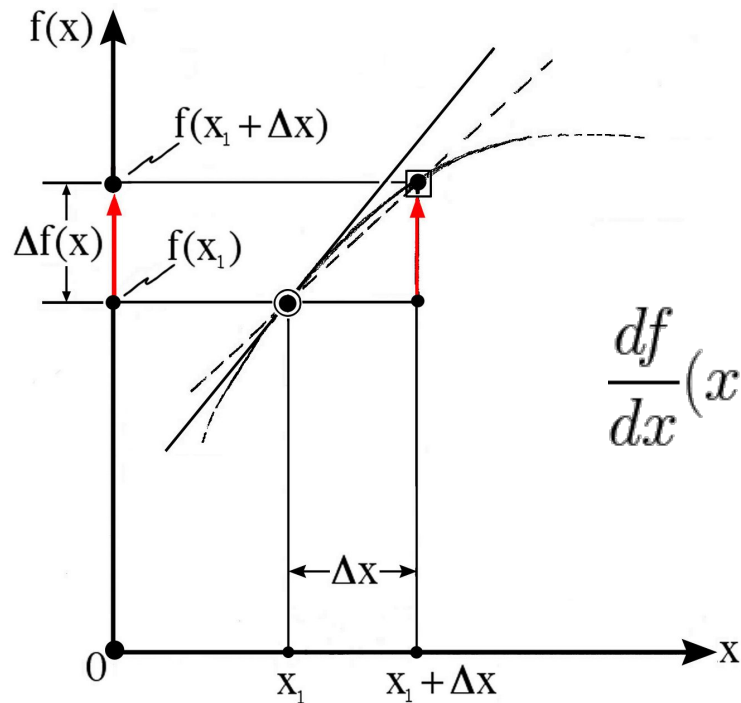
# Introdução ao problema de otimização

- Gradiente descendente (GD)



# Introdução ao problema de otimização

- Definição de gradiente (1D)



$$\frac{df}{dx}(x_1) = \lim_{\Delta x \rightarrow 0} \left( \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x} \right)$$

# Introdução ao problema de otimização

- Definição de gradiente (n-D)

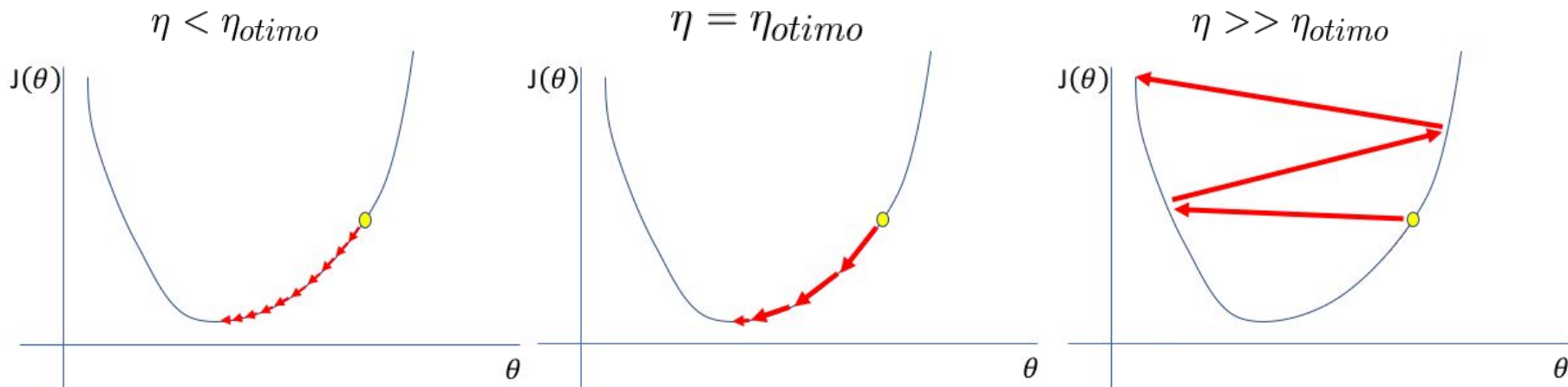
$$\mathbf{g} = \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$$

$$g_i = \frac{\partial}{\partial \theta_i} J(\boldsymbol{\theta})$$

# Introdução ao problema de otimização

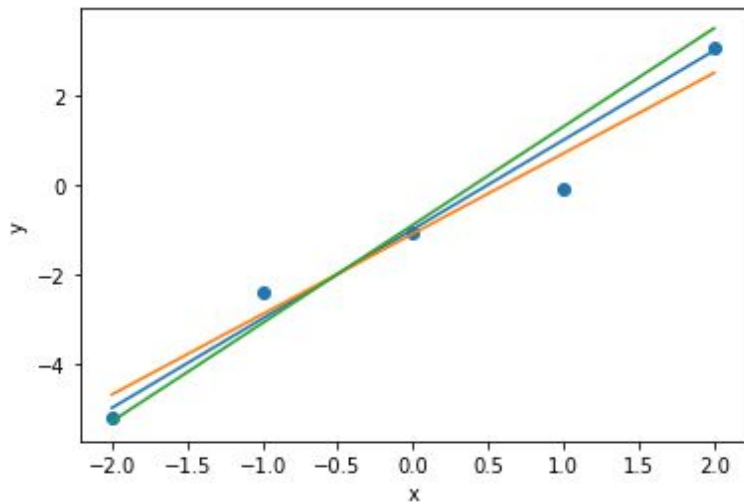
- Gradiente descendente (GD)

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta_t)$$



# Introdução ao problema de otimização

- Regressão linear!



$$\mathcal{L}(\Theta) = (\theta_0 x + \theta_1 - \hat{y})^2$$

$$\theta_{t_0} = np.random.rand((2,))$$

*while*  $\mathcal{L}(\Theta) > threshold$  :

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta)$$

$$\nabla_{\theta} \mathcal{L}(\theta) = \left[ \frac{\partial}{\partial \theta_0} \mathcal{L}(\theta), \frac{\partial}{\partial \theta_1} \mathcal{L}(\theta) \right]$$

# Introdução ao problema de classificação de imagens

- Problema da classificação de imagens



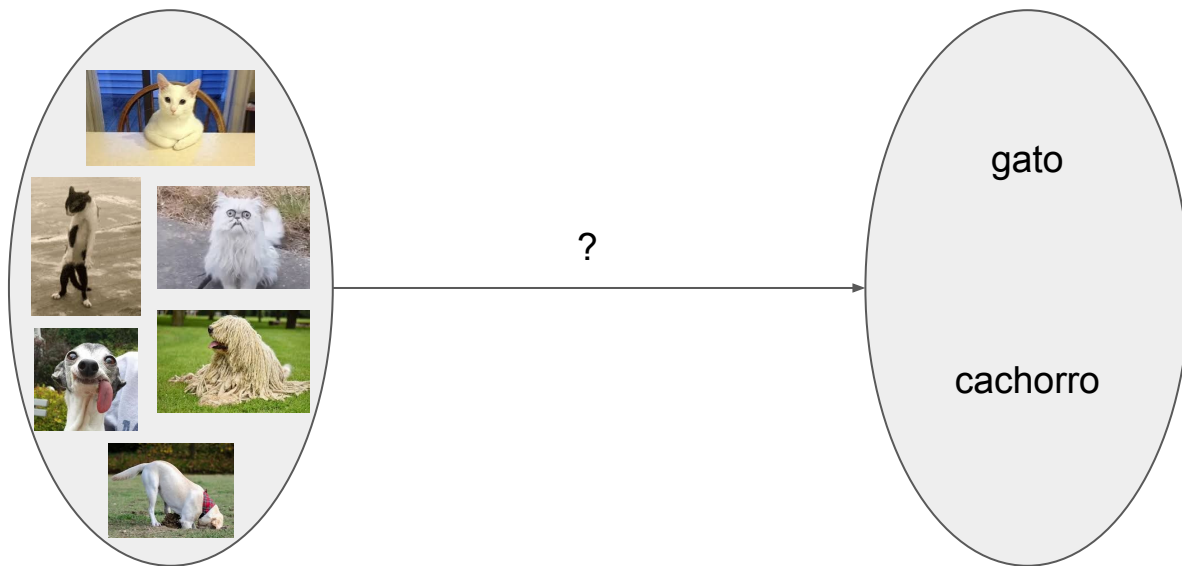
08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	87	28
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	48	01	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	57	48	30	03	49	13	36	65
52	70	95	23	04	60	11	42	69	21	65	96	01	32	56	71	37	02	36	91
22	31	16	71	51	62	63	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	15	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
55	44	65	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	35	25	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	54	65	69	82	67	59	55	74	04	36	16	
20	73	35	29	78	31	90	01	74	31	49	71	48	24	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	23	67	48

gato



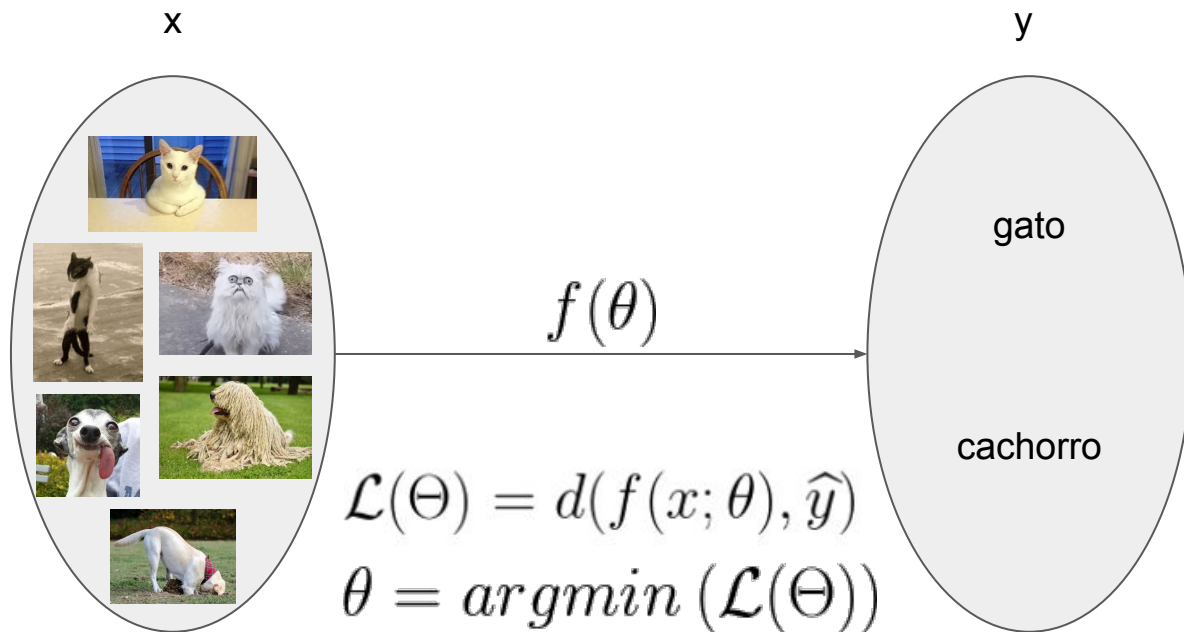
# Introdução ao problema de classificação de imagens

- Problema da classificação de imagens



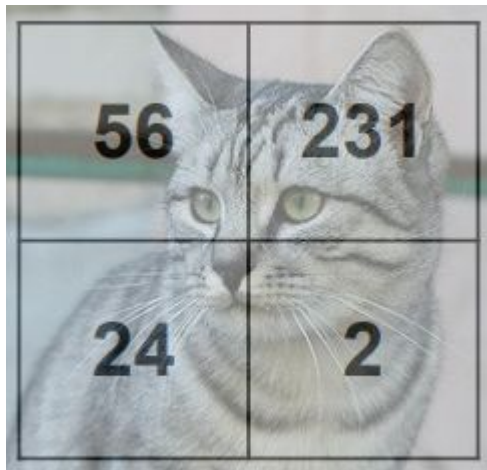
# Introdução ao problema de classificação de imagens

- Otimização para classificação de imagens



# Introdução ao problema de classificação de imagens

- Imagens como vetores



56
231
24
2

# Introdução ao problema de classificação de imagens

- Função linear

$\theta_{0,0}$	$\theta_{0,1}$	$\theta_{0,2}$	$\theta_{0,3}$
$\theta_{1,0}$	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{1,3}$

 $*$ 

<b>56</b>
<b>231</b>
<b>24</b>
<b>2</b>

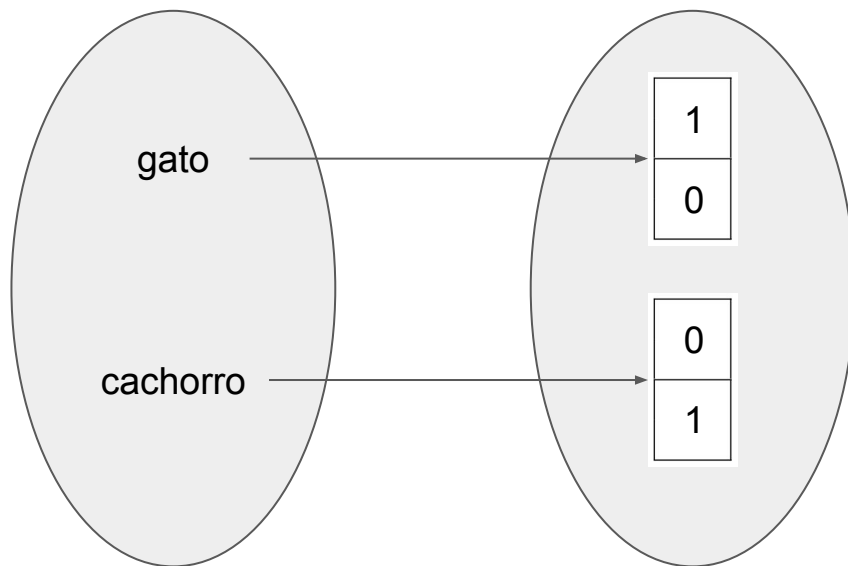
 $+$ 

$\theta_{0,4}$
$\theta_{1,4}$

 $=$   $f(x; \theta)$

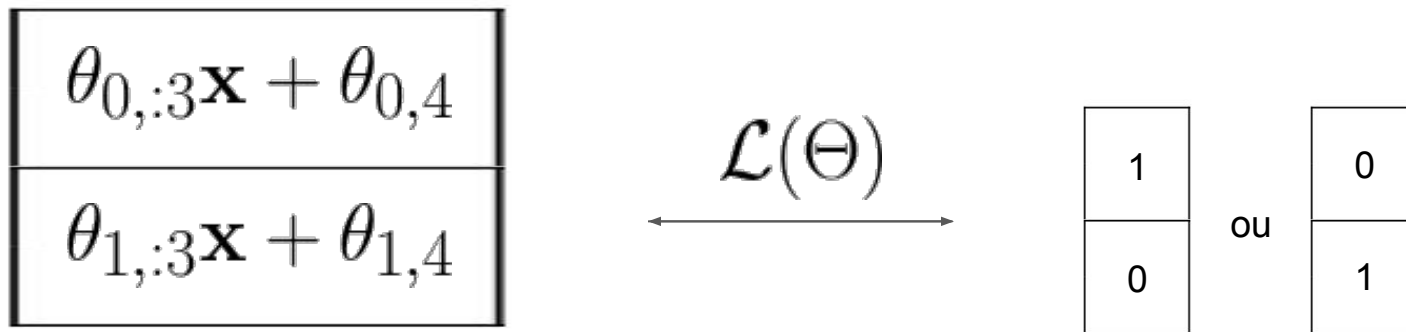
# Introdução ao problema de classificação de imagens

- Codificação one-hot da saída:



# Introdução ao problema de classificação de imagens

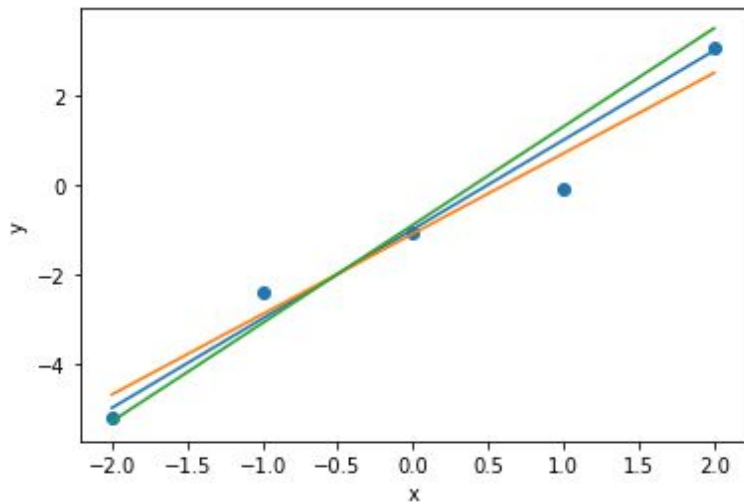
- Função custo



$$\mathcal{L}(\Theta) = d(f(x; \theta), \hat{y}) = \|f(x, \theta) - \hat{y}\|_2^2$$

# Introdução ao problema de classificação de imagens

- Gradiente descendente



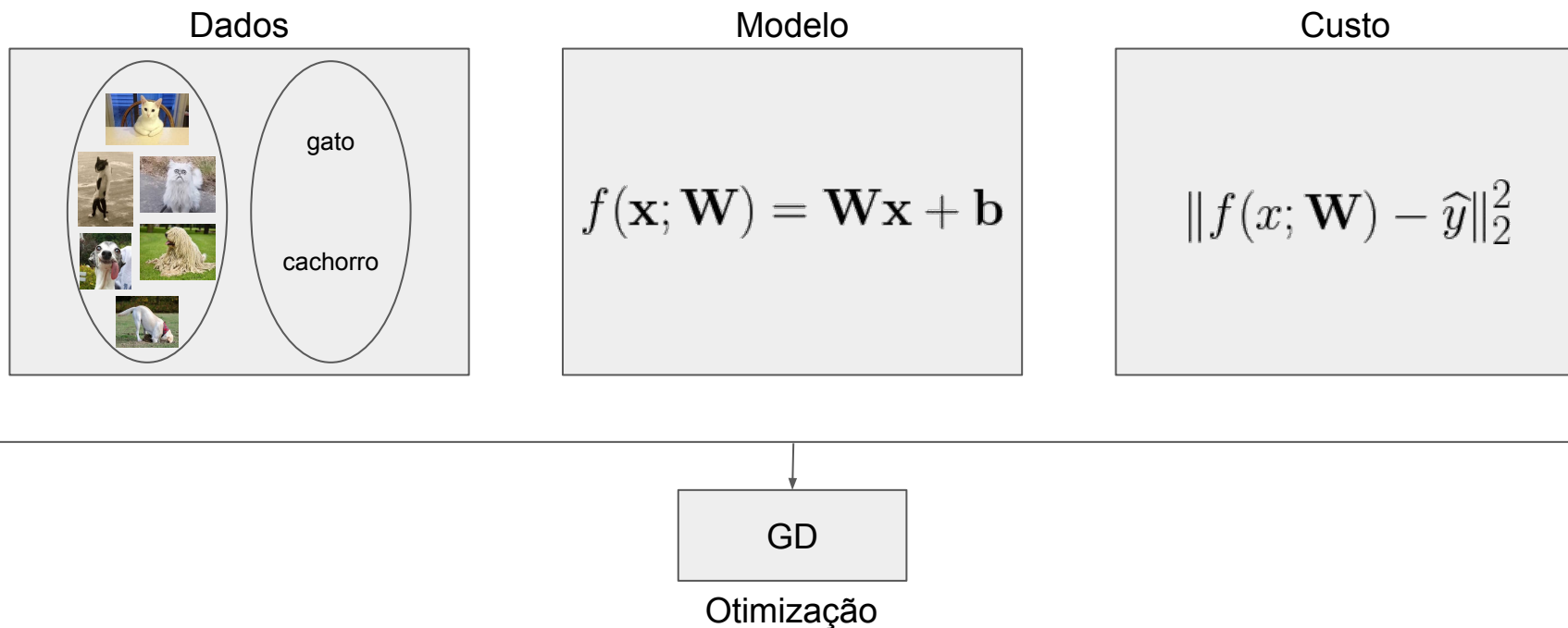
$\theta_{início} = np.random.rand((2, 4))$

*while*  $\mathcal{L}(\Theta) > threshold$  :

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}(\theta)$$

# Visão geral de problemas orientados a dados

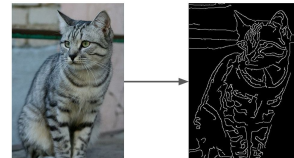
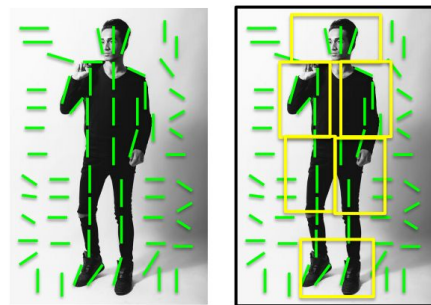
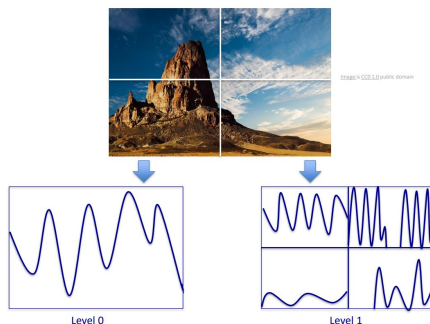
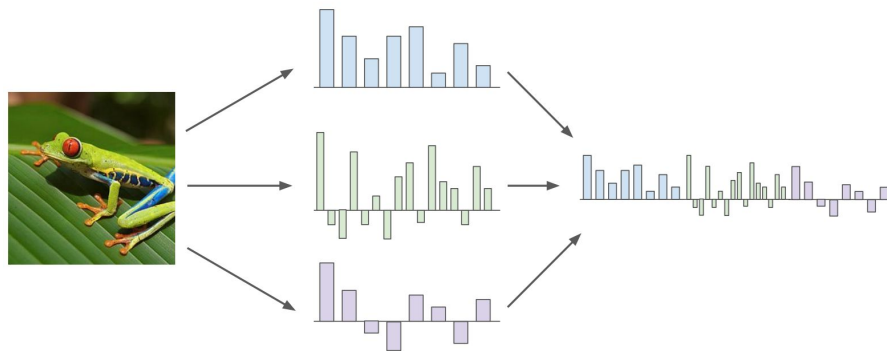
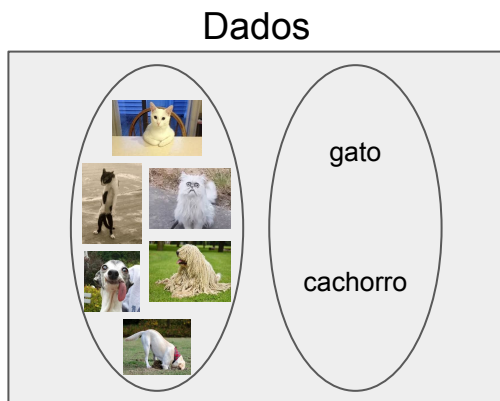
- Áreas para tomadas de decisões





# Visão geral de problemas orientados a dados

- Diferentes formas de representar os dados



# Visão geral de problemas orientados a dados

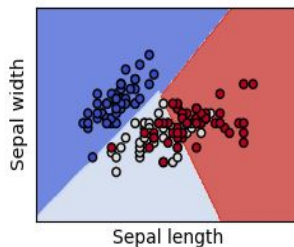
- Diferentes modelos

Modelo

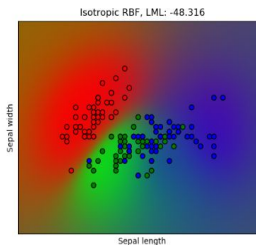
$$f(\mathbf{x}; \mathbf{W}) = \mathbf{W}\mathbf{x} + \mathbf{b}$$

SVM

SVC with linear kernel

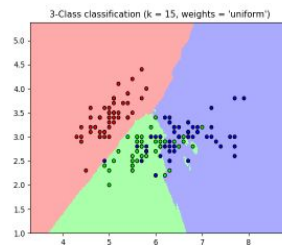


Mixture of Gaussians



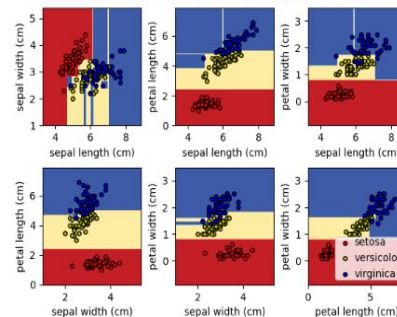
Naive Bayes

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i | y)$$

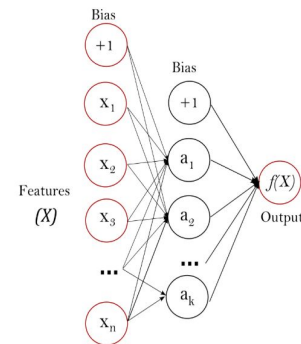


K-NN

Decision surface of a decision tree using paired features



Decision Trees



Redes Neurais

# Visão geral de problemas orientados a dados

- Diferentes funções custo

Custo

$$\|f(x; \mathbf{W}) - \hat{y}\|_2^2$$

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

$$H(p, q) = - \sum_x p(x) \log q(x)$$

...

$$d_p(x_i, x_j) = \|x_i - x_j\|_p = \left( \sum_n |x_{i_n} - x_{j_n}|^p \right)^{\frac{1}{p}}$$

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta, \\ \delta |y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases}$$

# Visão geral de problemas orientados a dados

- Diferentes algoritmos de otimização



SGD

SGD com momento

NAG

Adagrad

Adam

Radam

...

# Visão geral de problemas orientados a dados

- Detalhes nos próximos capítulos...

Os 300 (Hyper)Parâmetros

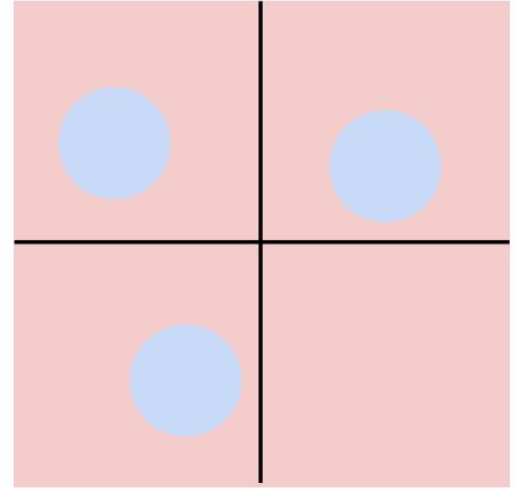
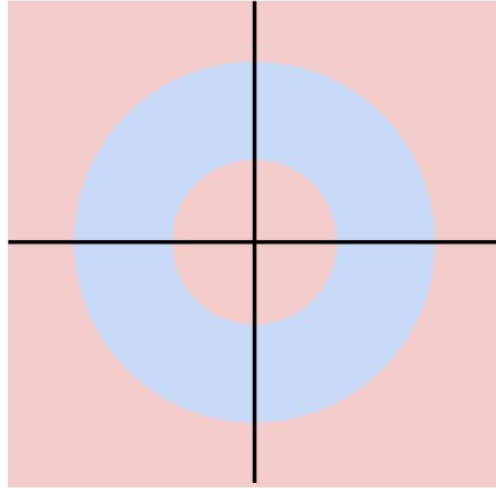
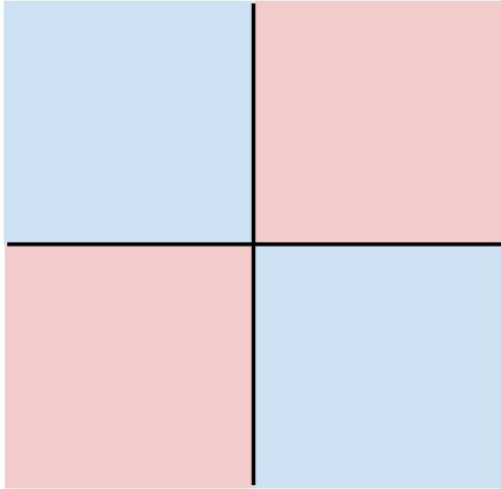


Lucas Araujo  
Instituto de Informática  
Universidade Federal de Goiás  
[www.deeplearningbrasil.com.br](http://www.deeplearningbrasil.com.br)



# Limitações de modelos lineares

- Modelos lineares não resolvem problemas não-lineares.... derr...



## No próximo episódio...

- Introdução a redes neurais
- Retropropagação (backpropagation)
- Introdução a redes convolucionais (CNN)

# Introdução ao problema de otimização

- Pontos críticos

$$\nabla_{\theta} J(\theta) = 0$$

