RNP-VC

Aula 1: Introdução a ML e VC

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

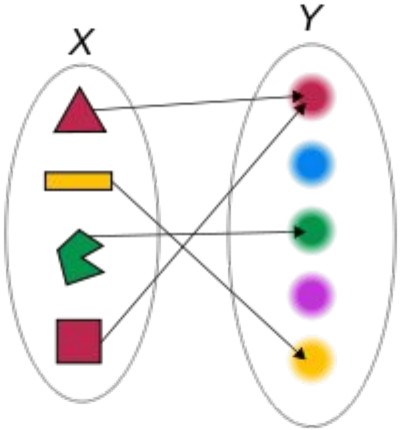
Sumário

- 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

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

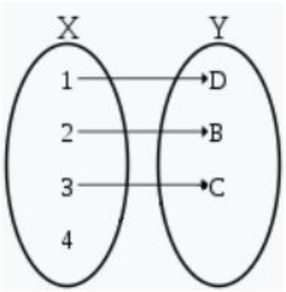


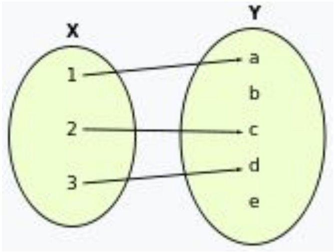
Introdução ao problema de otimização

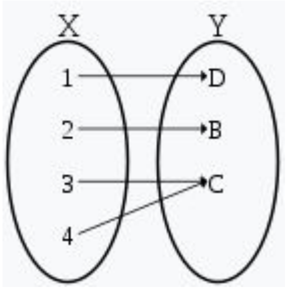
- Definição de função

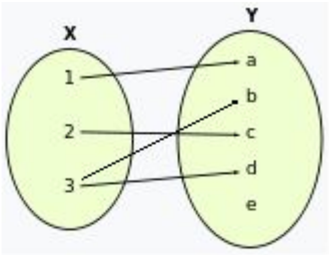
$$f : X \to Y, \mathbf{y} = f(\mathbf{x})$$

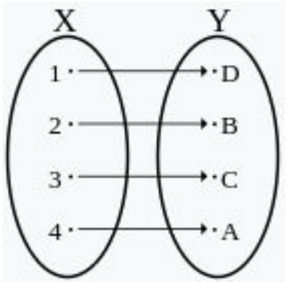
- É ou não é função? Por que?









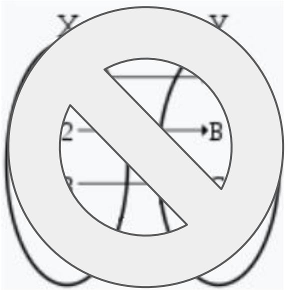


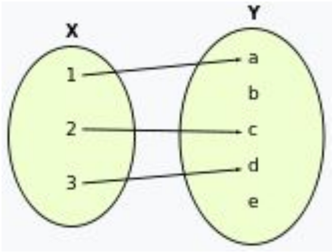
Introdução ao problema de otimização

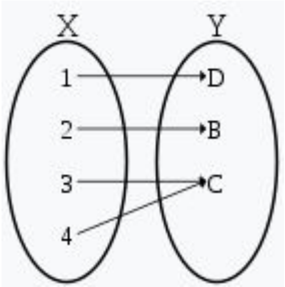
- Definição de função

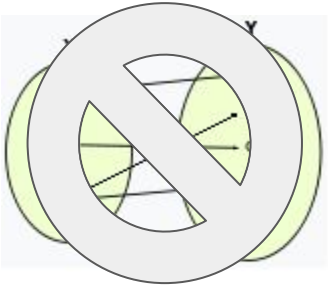
$$f : X \to Y, \mathbf{y} = f(\mathbf{x})$$

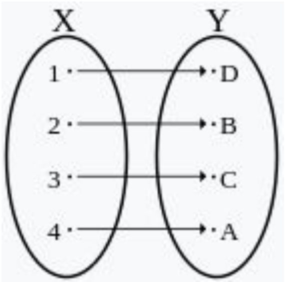
- É ou não é função? Por que?







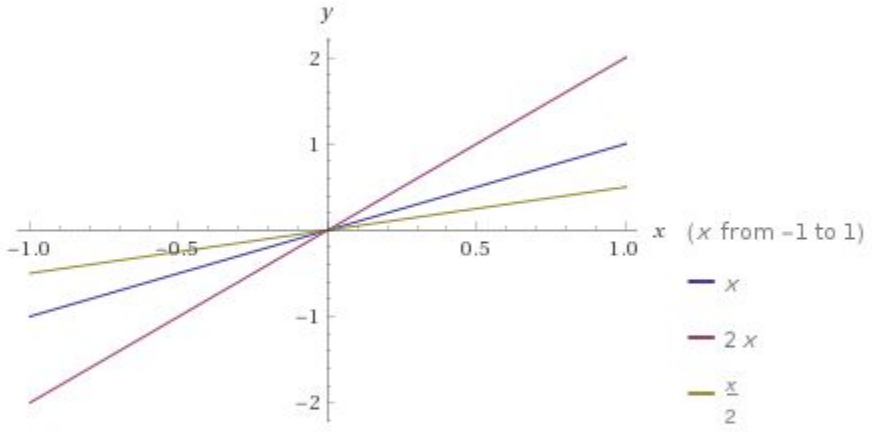




Introdução ao problema de otimização

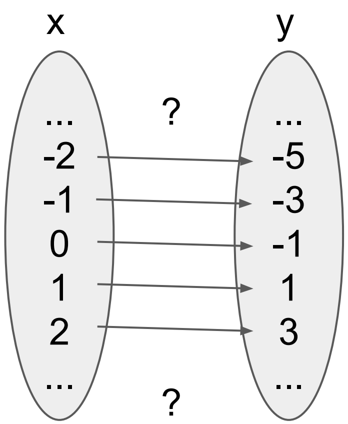
- Definição de função parametrizável

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



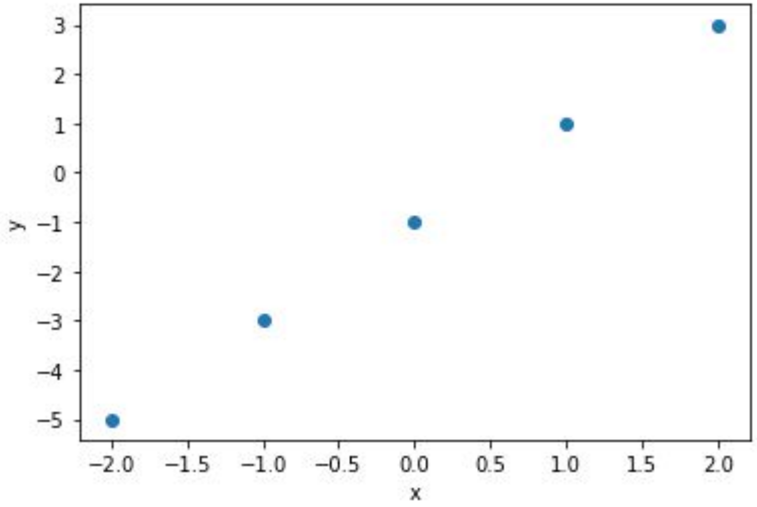
Introdução ao problema de otimização

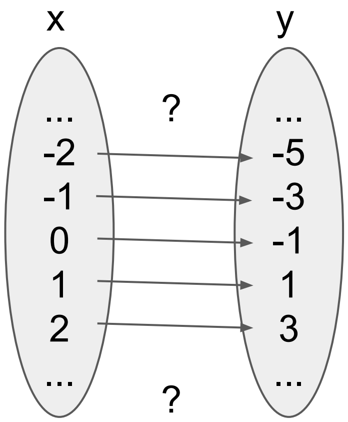
- Exemplo de dataset



Introdução ao problema de otimização

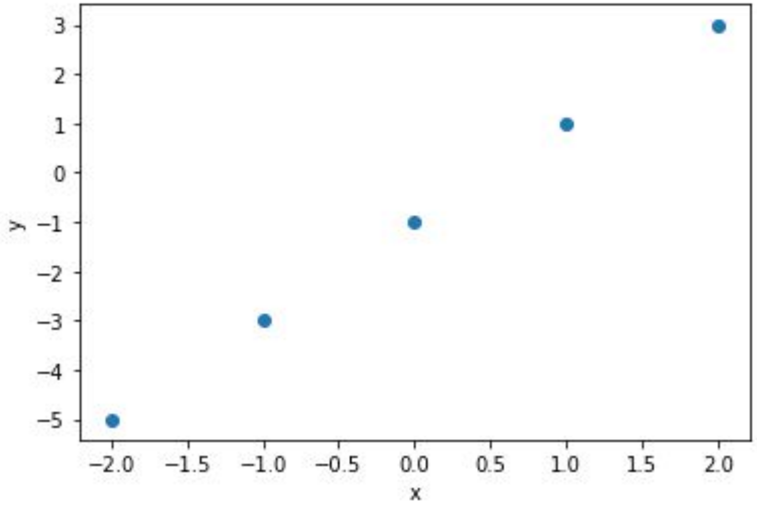
- Visualização dos dados

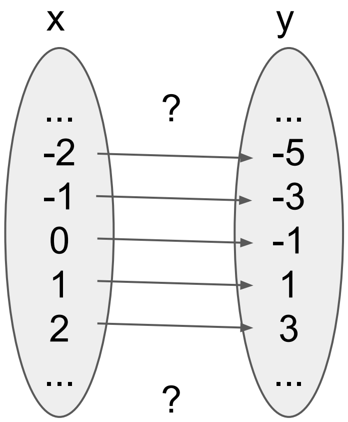




Introdução ao problema de otimização

- Regressão linear

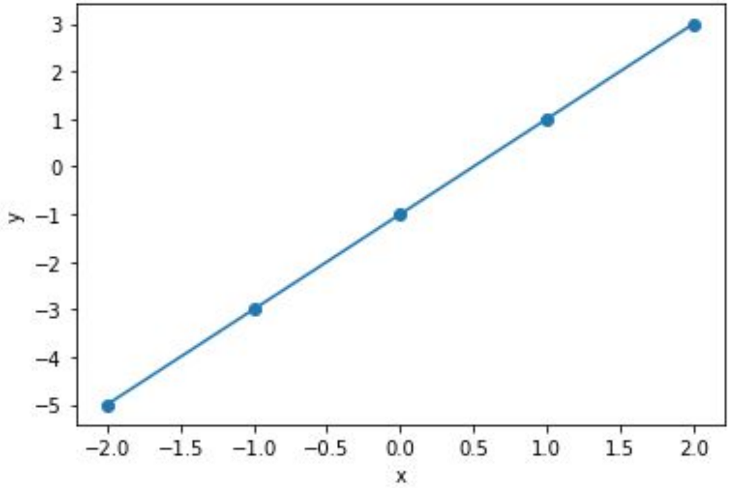




$$f : \mathbb{R}, \Theta \rightarrow \mathbb{R}, y = f(x;\theta) = \theta_0 x + \theta_1$$

Introdução ao problema de otimização

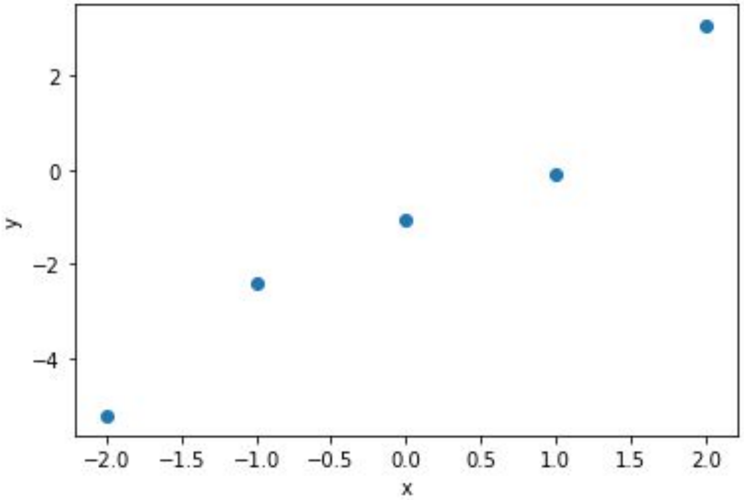
- Regressão linear



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

Introdução ao problema de otimização

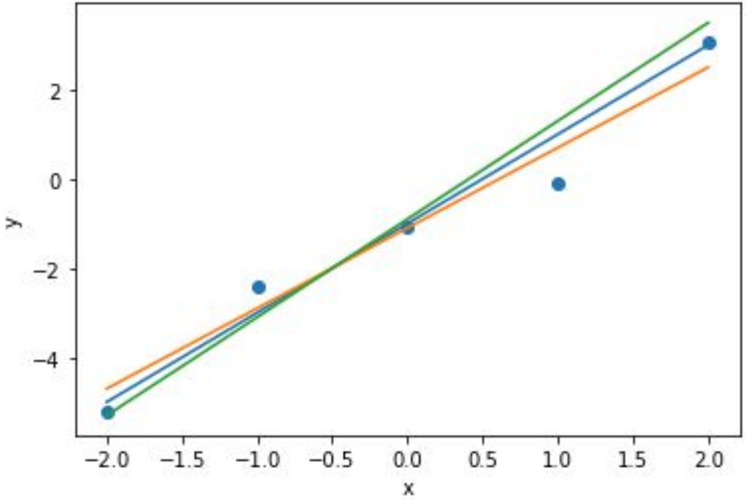
- Regressão linear?



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

Introdução ao problema de otimização

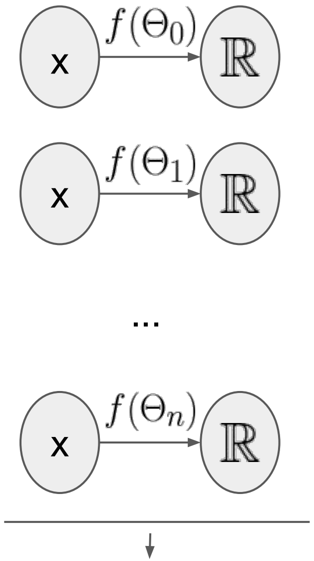
- Regressão linear?



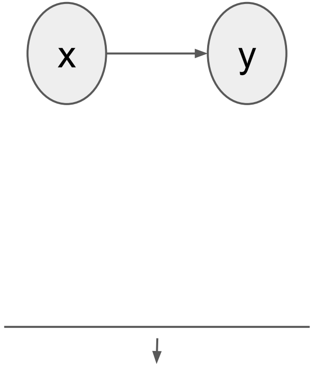
$$f:\mathbb{R},\Theta\to\mathbb{R}, y=f(x;\theta)=\theta_0 x+\theta_1$$

Introdução ao problema de otimização

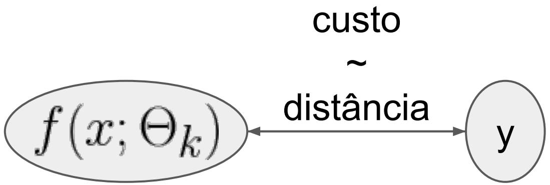
- Função custo (loss)



Conjunto de funções possíveis



Dados



Introdução ao problema de otimização

- Distância

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

$$1 : d(x_i, x_j) \geq 0$$
$$2 : d(x_i, x_j) = d(x_j, x_i)$$
$$3 : d(x_i, x_j) = 0 \Leftrightarrow x_i = x_j$$
$$4 : d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$$

Introdução ao problema de otimização

- Exemplos de distâncias

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

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

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

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

$$\vdots$$

$$d_\infty(x_i, x_j) = \left\| x_i - x_j \right\|_\infty = \underset{max}{n} \left| x_{i_n} - x_{j_n} \right|$$

Introdução ao problema de otimização

- Custo

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

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

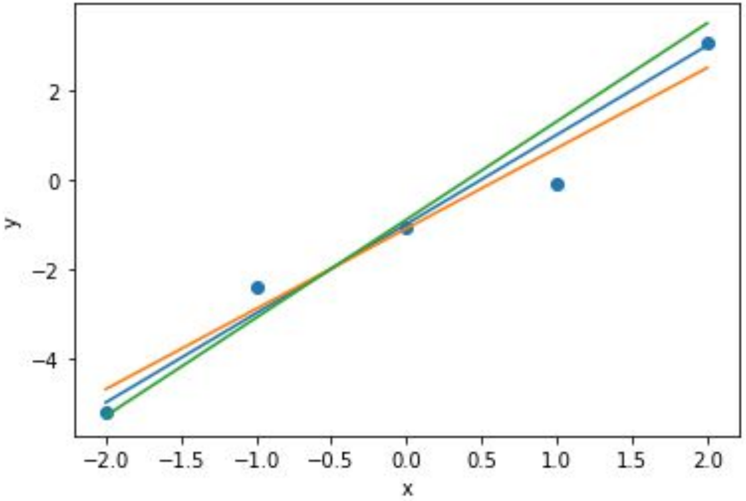
$$\cancel{2 : d(x_i, x_j) = d(x_j, x_i)}$$

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

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

Introdução ao problema de otimização

- Regressão linear!



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

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

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

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

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

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

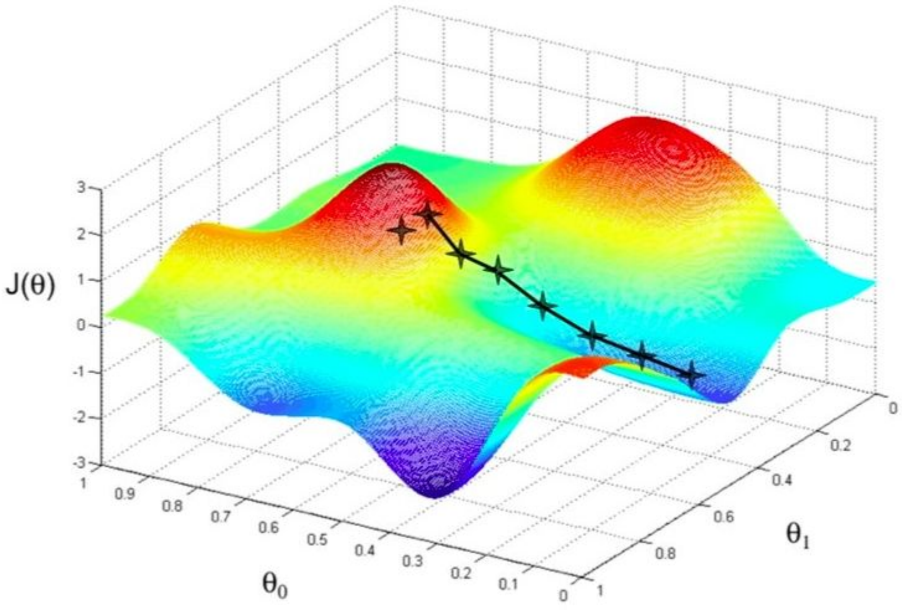
Introdução ao problema de otimização

- Otimização: como achar o theta?

    - Solução analítica do problema
    - Busca exaustiva
    - Busca aleatória (ex: algoritmos genéticos)
    - Busca baseada em gradientes (ex: gradiente descendente)

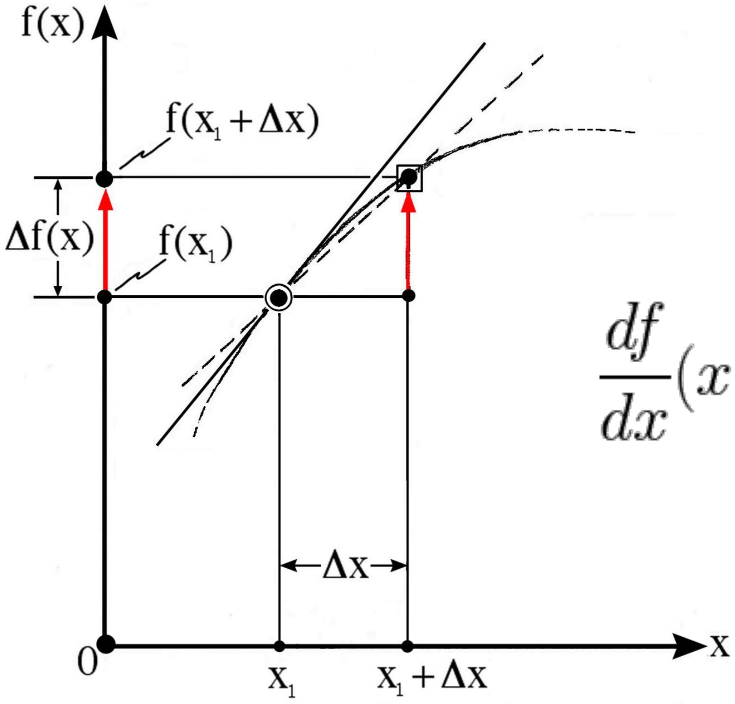
Introdução ao problema de otimização

- Gradiente descendente (GD)



Introdução ao problema de otimização

- Definição de gradiente (1D)



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

Introdução ao problema de otimização

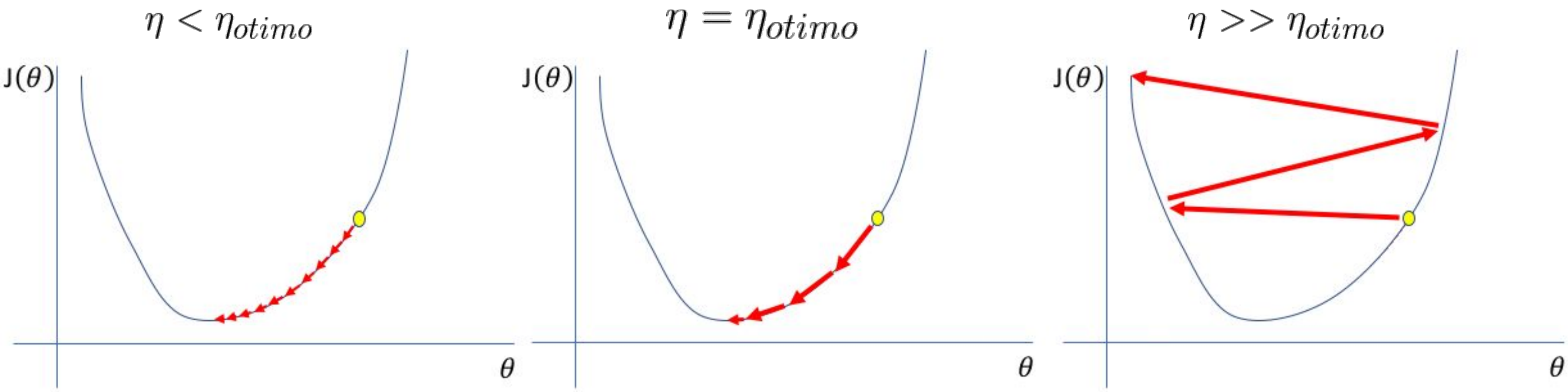
- Definição de gradiente (n-D)

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

Introdução ao problema de otimização

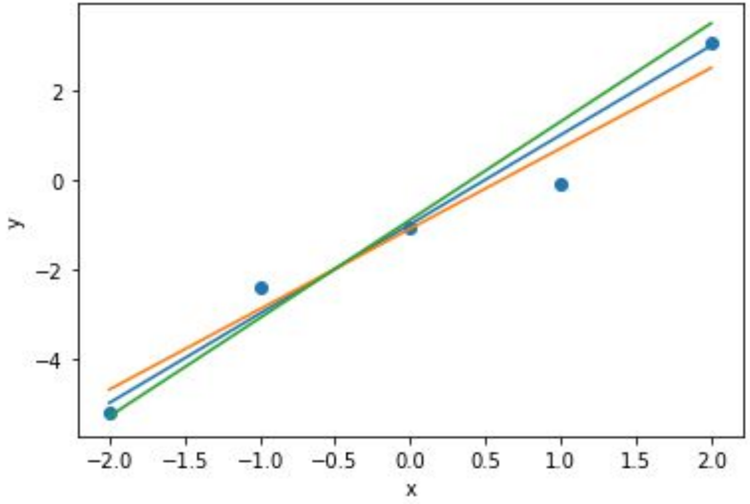
- Gradiente descendente (GD)

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



Introdução ao problema de otimização

- Regressão linear!



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

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

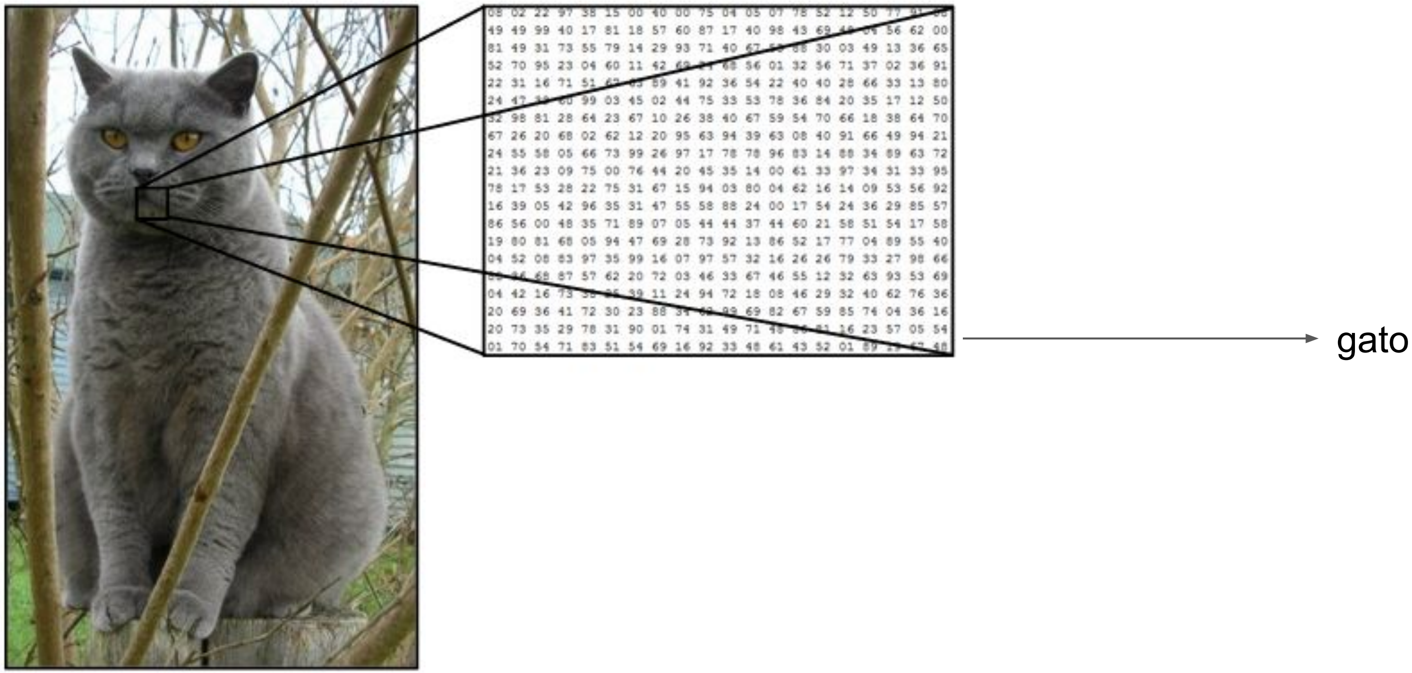
$$while\ \mathcal{L}(\Theta) > threshold:$$

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

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

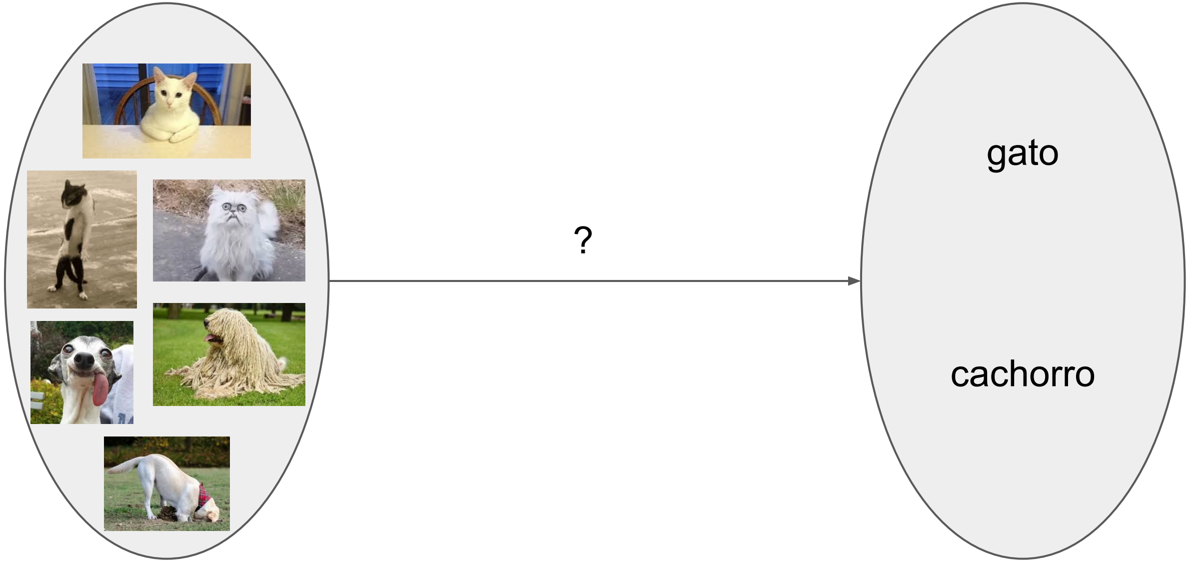
Introdução ao problema de classificação de imagens

- Problema da classificação de imagens



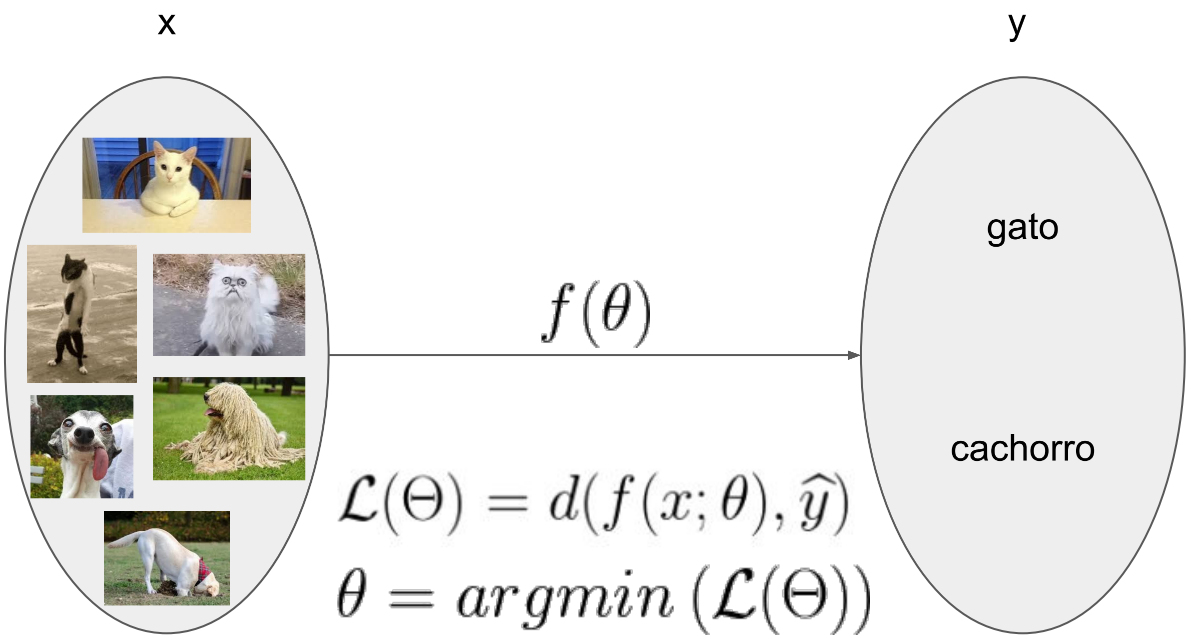
Introdução ao problema de classificação de imagens

- Problema da classificação de imagens



Introdução ao problema de classificação de imagens

- Otimização para classificação de imagens

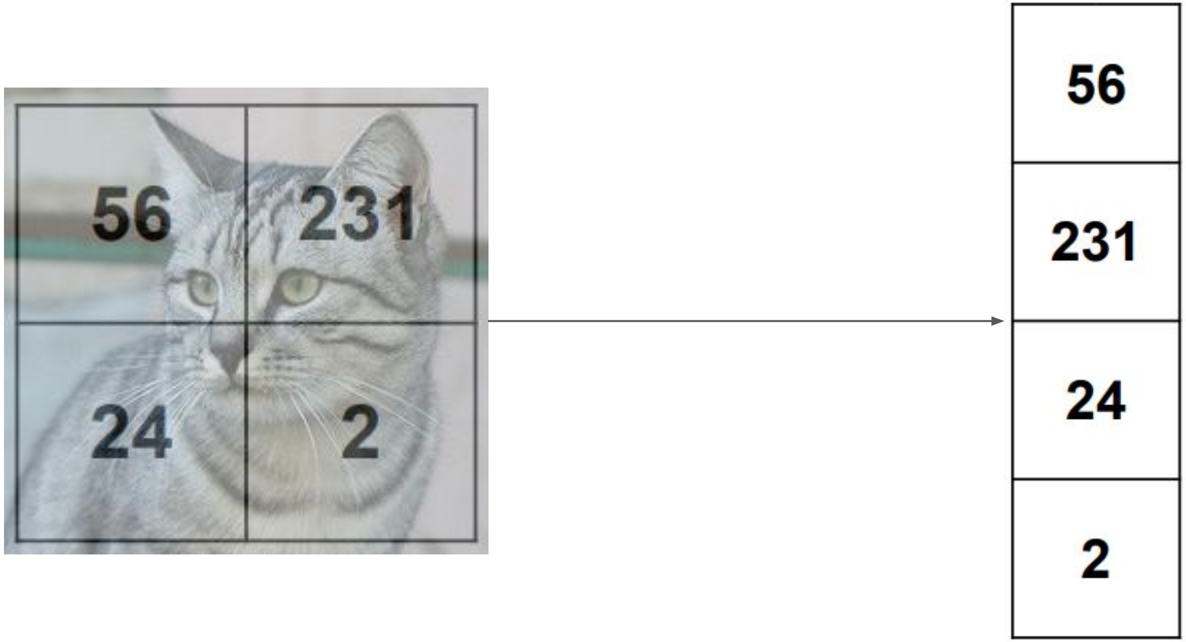


$$\mathcal{L}(\Theta) = d(f(x;\theta), \widehat{y})$$

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

Introdução ao problema de classificação de imagens

- Imagens como vetores



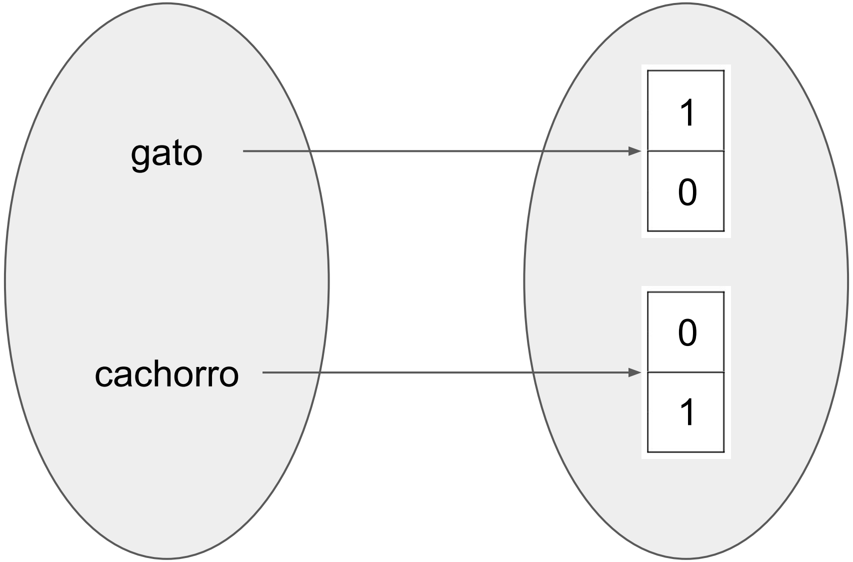
Introdução ao problema de classificação de imagens

- Função linear

$$
\begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \theta_{0,2} & \theta_{0,3} \\ \theta_{1,0} & \theta_{1,1} & \theta_{1,2} & \theta_{1,3} \end{bmatrix} * \begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix} + \begin{bmatrix} \theta_{0,4} \\ \theta_{1,4} \end{bmatrix} = f(x;\theta)
$$

Introdução ao problema de classificação de imagens

- Codificação one-hot da saída:



Introdução ao problema de classificação de imagens

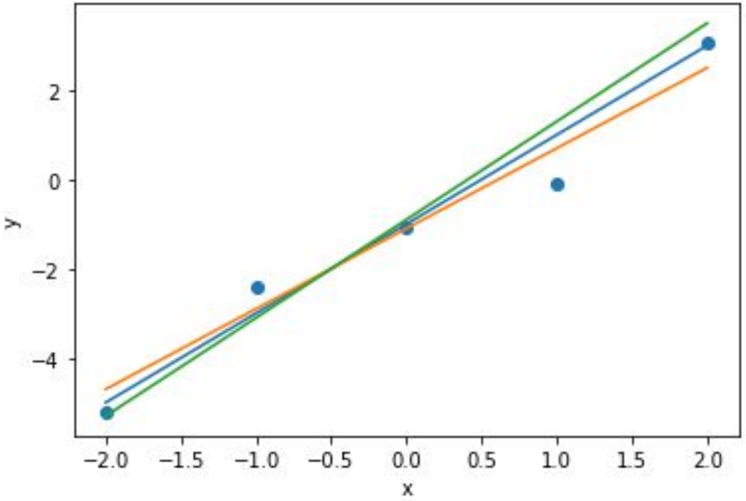
- Função custo

$$\begin{bmatrix} \theta_{0,:3}\mathbf{x} + \theta_{0,4} \\ \theta_{1,:3}\mathbf{x} + \theta_{1,4} \end{bmatrix} \xleftrightarrow{\mathcal{L}(\Theta)} \begin{bmatrix}1\\0\end{bmatrix} \text{ ou } \begin{bmatrix}0\\1\end{bmatrix}$$

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

Introdução ao problema de classificação de imagens

- Gradiente descendente



$$\theta_{inicio} = np.random.rand((2,4))$$
$$while\ \mathcal{L}(\Theta) > threshold:$$
$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\theta)$$

Visão geral de problemas orientados a dados

- Áreas para tomadas de decisões

Dados

gato

cachorro

Modelo

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

Custo

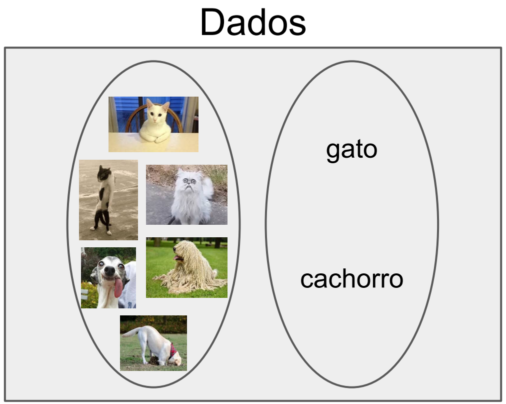
$$\|f(x; \mathbf{W}) - \widehat{y}\|_2^2$$

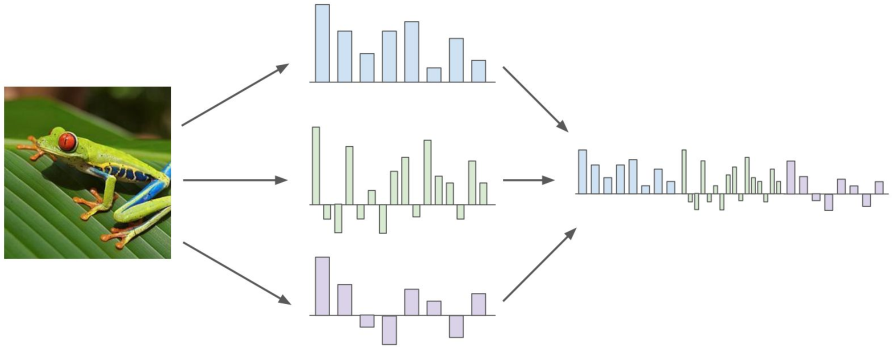
GD

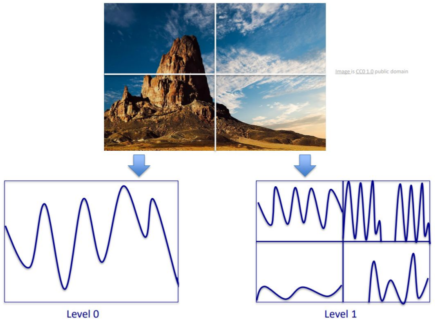
Otimização

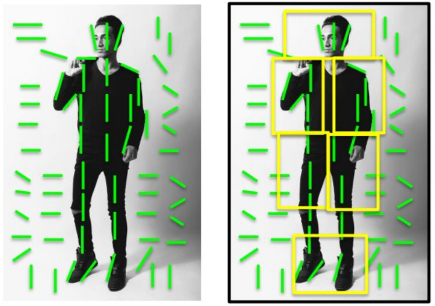
Visão geral de problemas orientados a dados

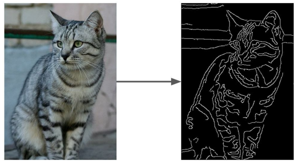
- Diferentes formas de representar os dados











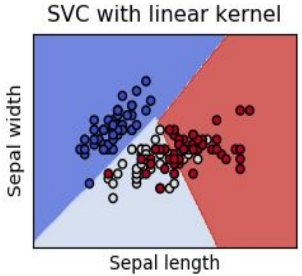
Visão geral de problemas orientados a dados

- Diferentes modelos

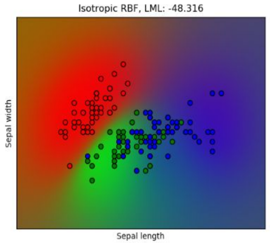
Modelo

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

SVM

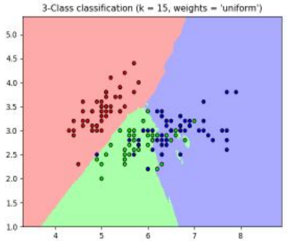


Mixture of Gaussians

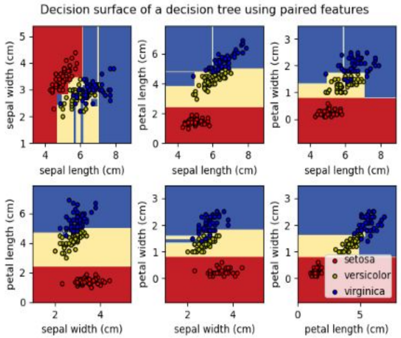


Naive Bayes

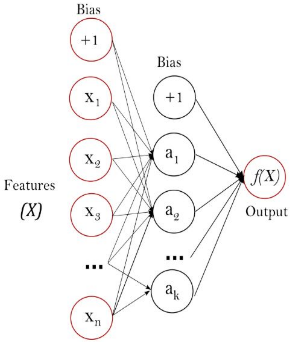
$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y)$$



K-NN



Decision Trees



Redes Neurais

...

Visão geral de problemas orientados a dados

- Diferentes funções custo

Custo

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

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

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

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

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

...

Visão geral de problemas orientados a dados

- Diferentes algoritmos de otimização

Otimização

GD

SGD

SGD com momento

NAG

Adagrad

Adam

Radam

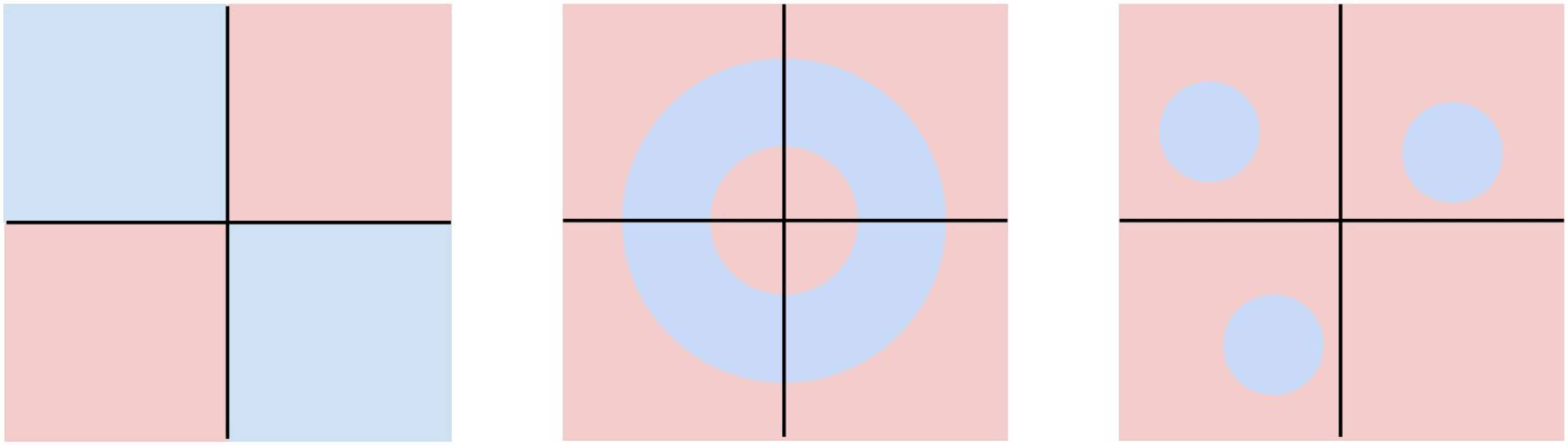
...

Visão geral de problemas orientados a dados

- Detalhes nos próximos capítulos...

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

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