



**Universidade Federal da Paraíba**

Coordenação do Curso de Ciência de Dados e  
Inteligência Artificial



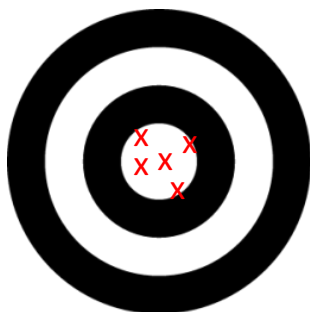
## **Compromisso Viés-Variância e Overfitting**

Prof. Gilberto Farias

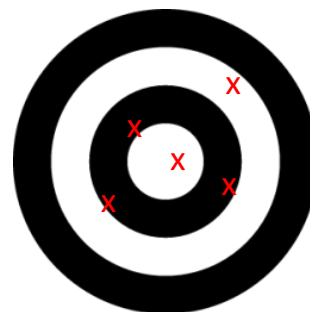
# Roteiro

- Tipos de erro
  - Como eliminar ruído da amostra?
  - Como eliminar viés da amostra?
- Compromisso Viés-Variância
- Overfitting
  - Definição
  - Regularização
  - Prática 1 : Heurística Weight Decay

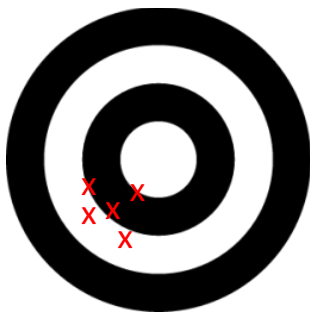
$$erro = viés + ruído$$



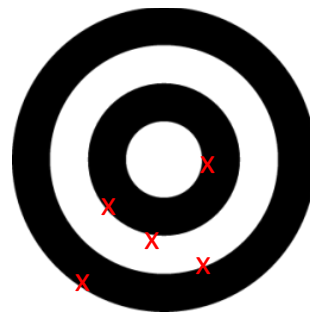
boa precisão



ruído



viés



viés e ruído

# Como identificar os erros no julgamento humano??



A cluster of 6 red 'x' marks arranged in a tight, roughly circular pattern, representing data points with low variance.



A cluster of 6 red 'x' marks arranged in a loose, irregular pattern, representing data points with high variance.

Ruído é identificado pela variância dos dados

Viés precisa do alvo para ser identificado

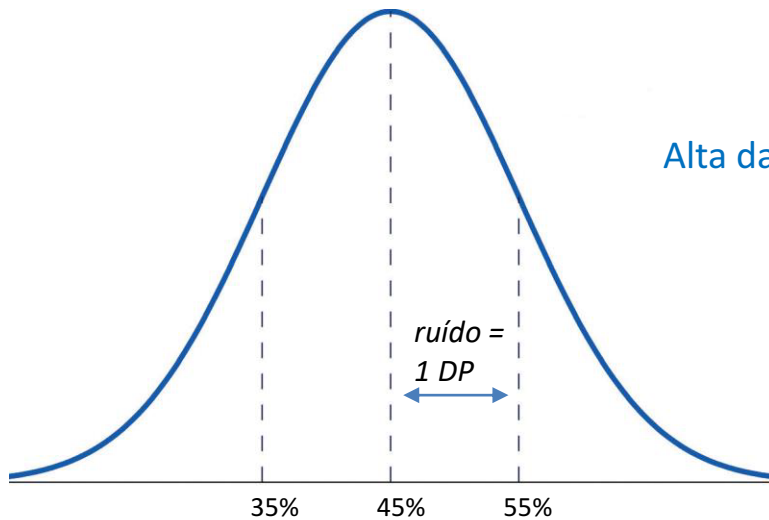


A cluster of 6 red 'x' marks arranged in a tight, roughly circular pattern, representing data points with low variance.



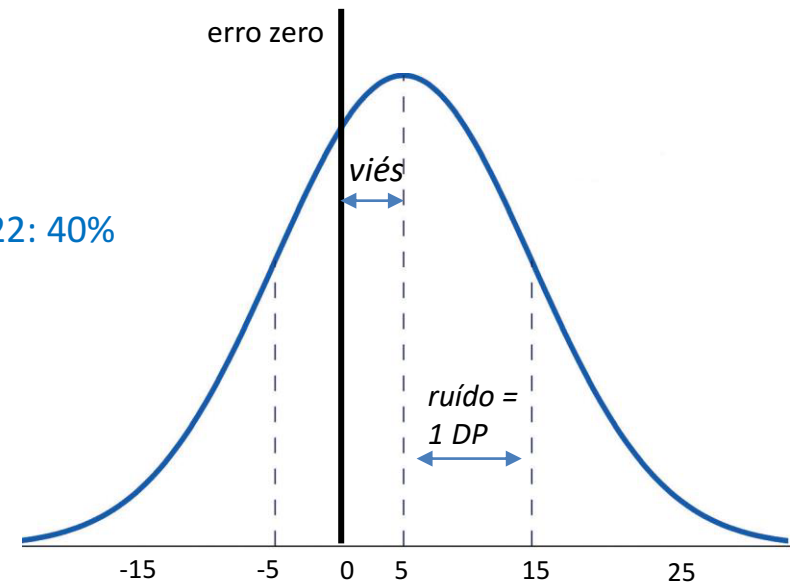
A cluster of 6 red 'x' marks arranged in a loose, irregular pattern, representing data points with high variance.

# Como computar o viés de uma previsão/julgamento?



Previsão da alta na bolsa de valores no ano de 2022 por especialistas

Alvo  
Alta da bolsa em 2022: 40%



Distribuição dos erros nas previsões dos especialistas

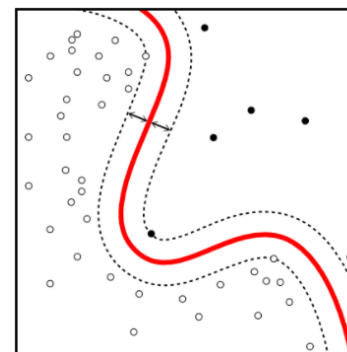
Como eliminar **viés** da amostra?

Uma amostra de classificações/julgamentos não representa o alvo.

Não tem como eliminar o viés da amostra

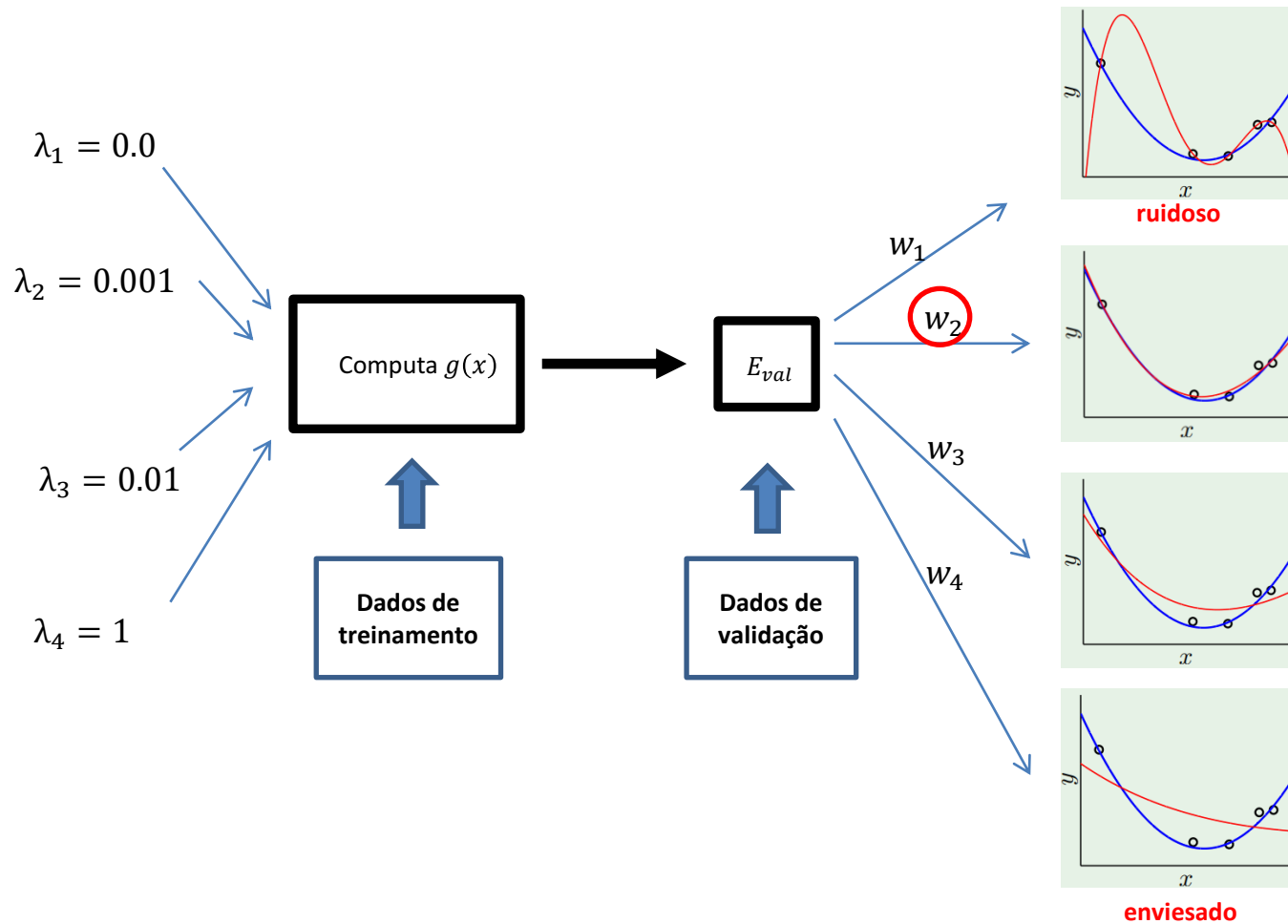
# E os erros no julgamento das funções inferidas dos dados??

Aprendizado  
de máquina



dataset

# Como eliminar **ruído** no Aprendizado Supervisionado?





# Compromisso Viés-Variância

# Compromisso de aproximação - generalização

- Queremos  $E_{out}$  pequeno  $\rightarrow$  boa aproximação de  $f$  fora da amostra;
  - $H$  complexo  $\rightarrow$  melhor chance de aproximar de  $f$ ;
  - $H$  mais complexo  $\rightarrow$  maior chance de generalizar fora da amostra;
- Mais fácil ganhar na loteria!
- Ideal  $H = \{f\}$

## Quantificando o compromisso

- A análise VC apresenta um limite de generalização de aprendizado:

$$E_{out} \leq E_{in} + \Omega$$

- Um conjunto  $H$  simples pode falhar na aproximação a função alvo  $f$
- Um conjunto  $H$  complexo pode falhar na generalização devido a seus termos complexos (aprender ruído)

$E_{in} \uparrow$

$E_{out} \uparrow$

# Viés - Variância

- A análise viés-variância decompõe o  $E_{out}$  em:
  - Quão bem  $H$  pode se aproximar de  $f$
  - Quão bem podemos encontrar um bom  $h \in H$
- Para a análise viés – variância tanto o conjunto de  $H$  como o algoritmo de aprendizagem  $A$  importam:

$g^{(D)}$  da regressão linear é diferente da  
 $g^{(D)}$  da regressão logística

- Análise viés-variância não pode ser implementada pois depende de  $f$ . É uma ferramenta conceitual utilizada na regularização

# Desmembrando $E_{out}$

$$E_{out} = \mathbb{E}_{\mathbf{x}} \left[ \left( g^{(D)}(\mathbf{x}) - f(\mathbf{x}) \right)^2 \right]$$



$$vies(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} [(\bar{g}(\mathbf{x}) - f(\mathbf{x}))^2]$$

$$var(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[ \left( g^{(D)}(\mathbf{x}) - \bar{g}(\mathbf{x}) \right)^2 \right]$$

$\bar{g}(\mathbf{x})$  : é a função média entre todos os possíveis conjuntos de dados  $D_1 \dots D_k$

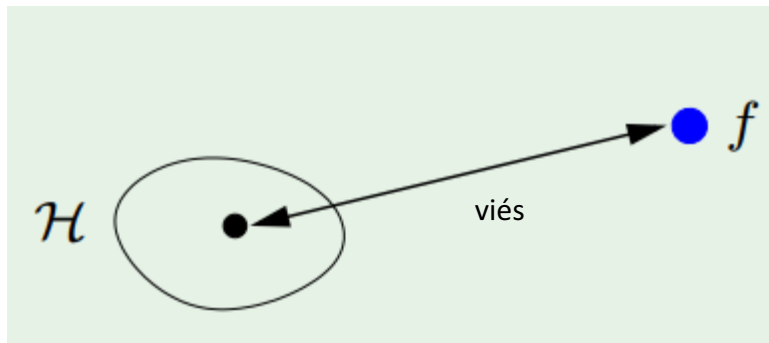
# Desmembrando $E_{out}$

$$E_{out} = \mathbb{E}_{\mathbf{x}} \left[ \left( g^{(D)}(\mathbf{x}) - f(\mathbf{x}) \right)^2 \right]$$



$$vies(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[ \left( \bar{g}(\mathbf{x}) - f(\mathbf{x}) \right)^2 \right]$$

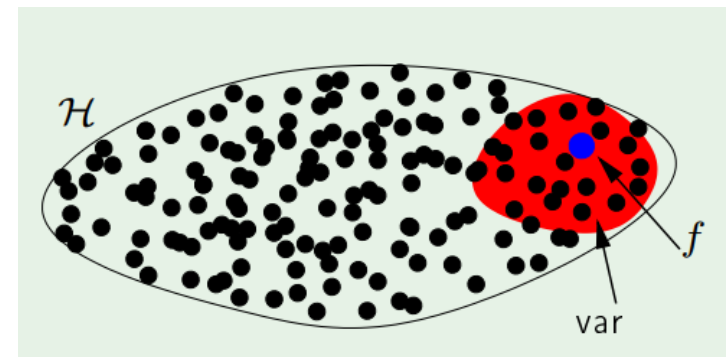
$$var(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[ \left( g^{(D)}(\mathbf{x}) - \bar{g}(\mathbf{x}) \right)^2 \right]$$



$$g^{(D)}(\mathbf{x}) = \bar{g}(\mathbf{x})$$

$$var(\mathbf{x}) = 0$$

$vies(\mathbf{x})$  grande



$\bar{g}(\mathbf{x})$  próximo de  $f$

$$vies(\mathbf{x}) \approx 0$$

$var(\mathbf{x})$  grande

# Exemplo: função alvo $\sin(x)$

$$f: [-1, +1] \rightarrow \mathbb{R} \quad f(x) = \sin(\pi x)$$

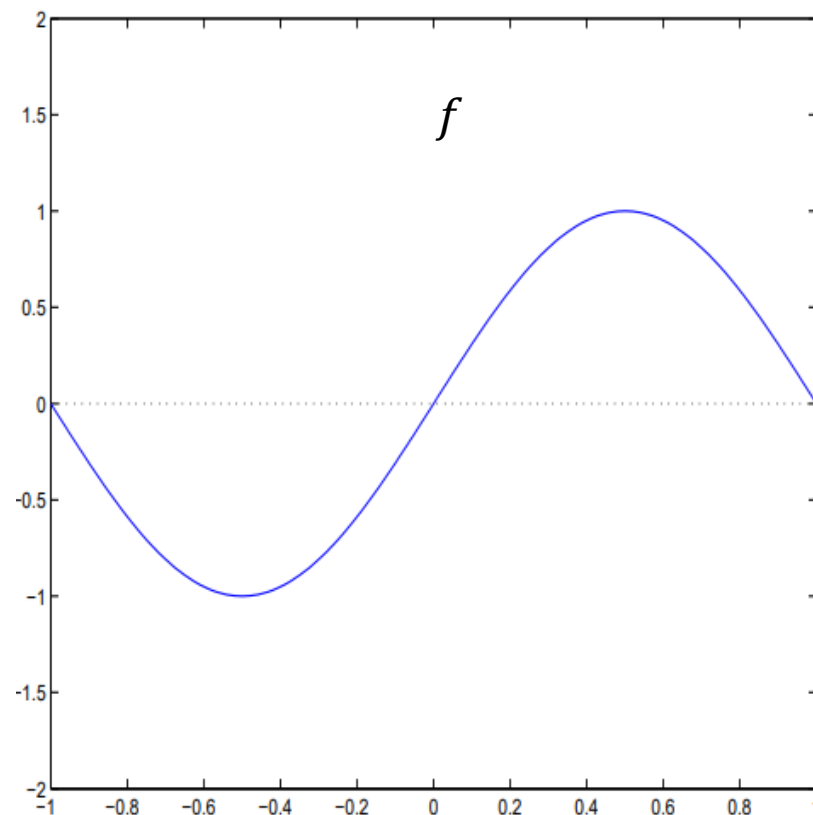
Duas amostras apenas!  $N = 2$

Modelos de Aprendizagem utilizados:

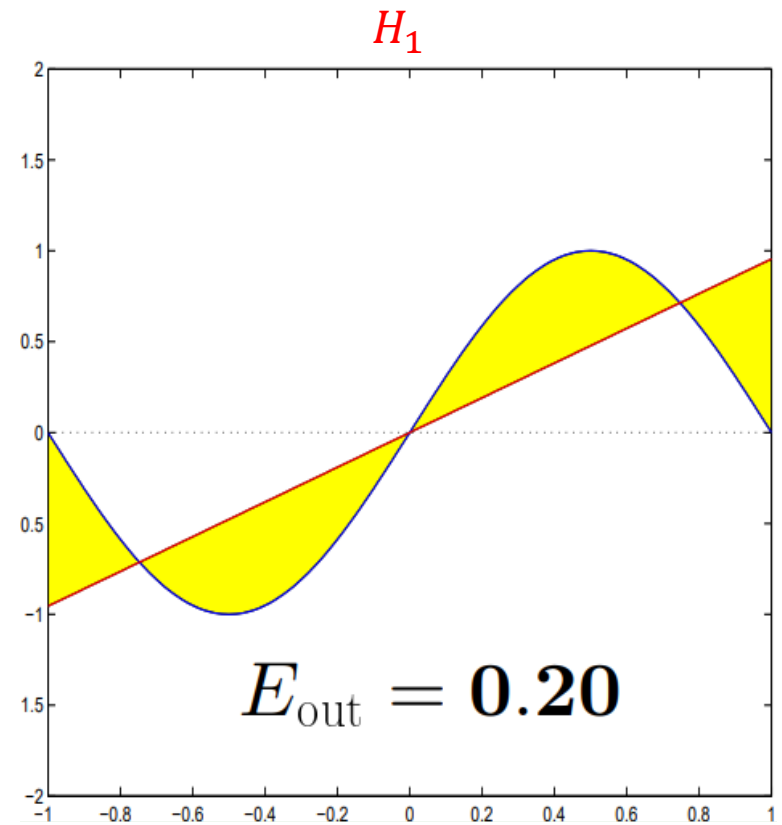
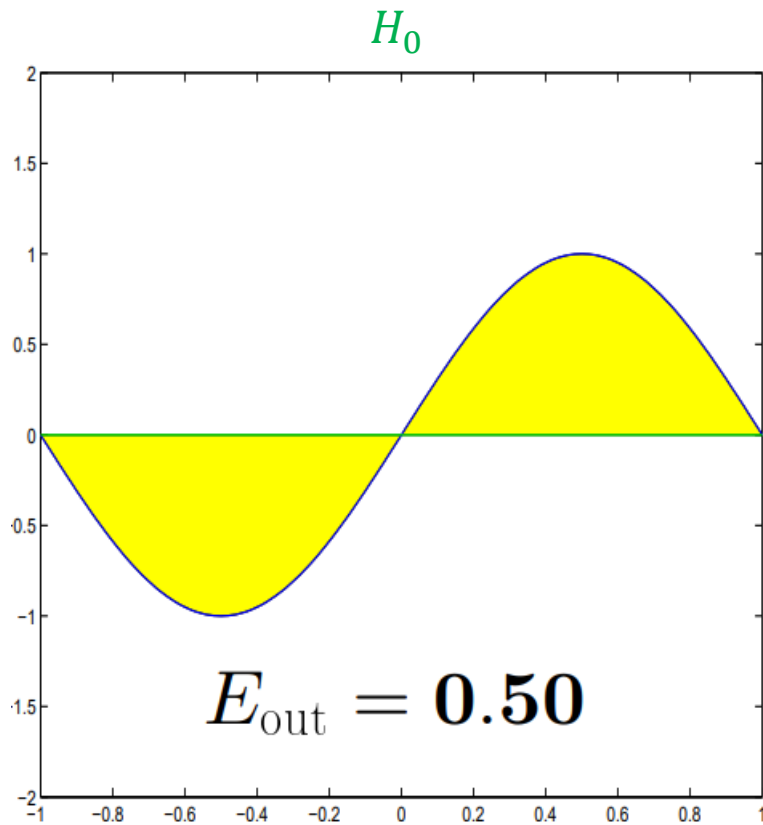
$$H_0: h(x) = b$$

$$H_1: h(x) = ax + b$$

Quem é o melhor  $H_0$  ou  $H_1$ ?

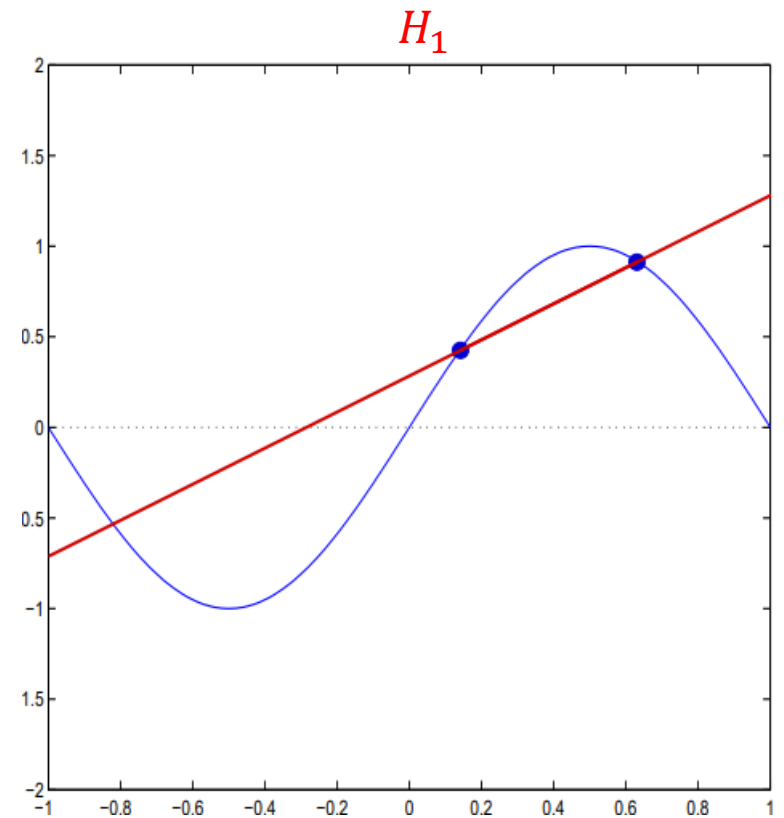
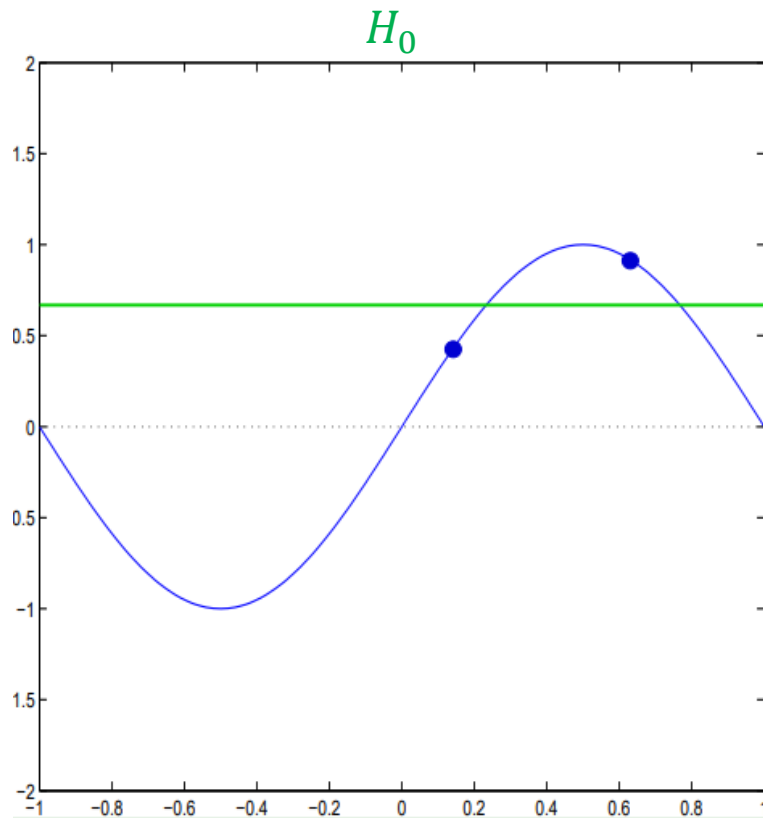


# Aproximação - $H_0 \times H_1$

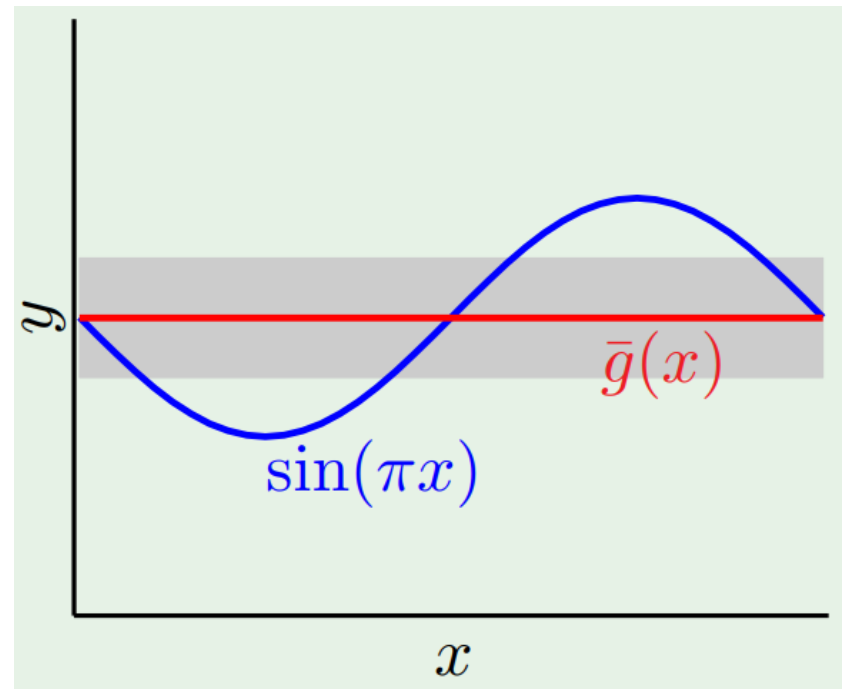
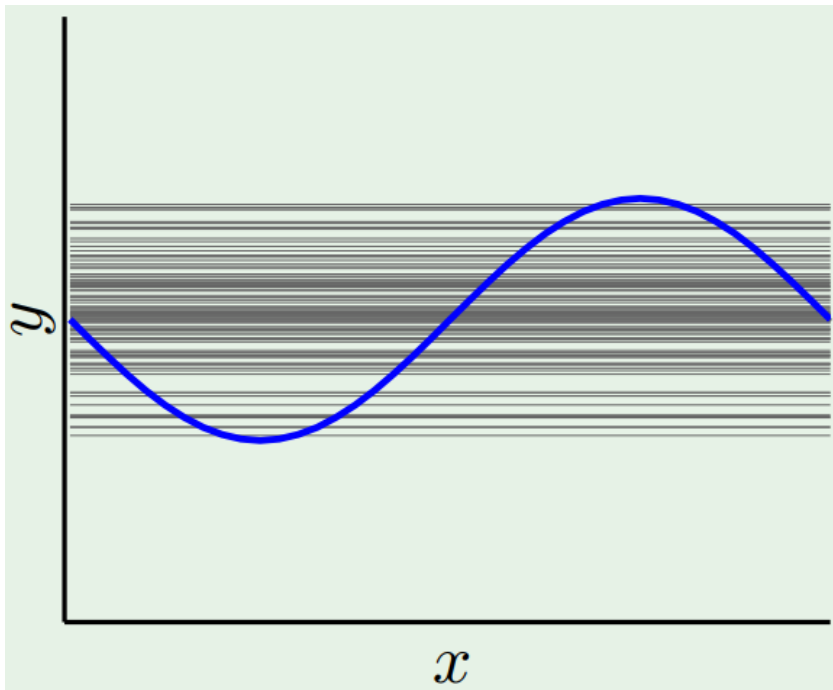




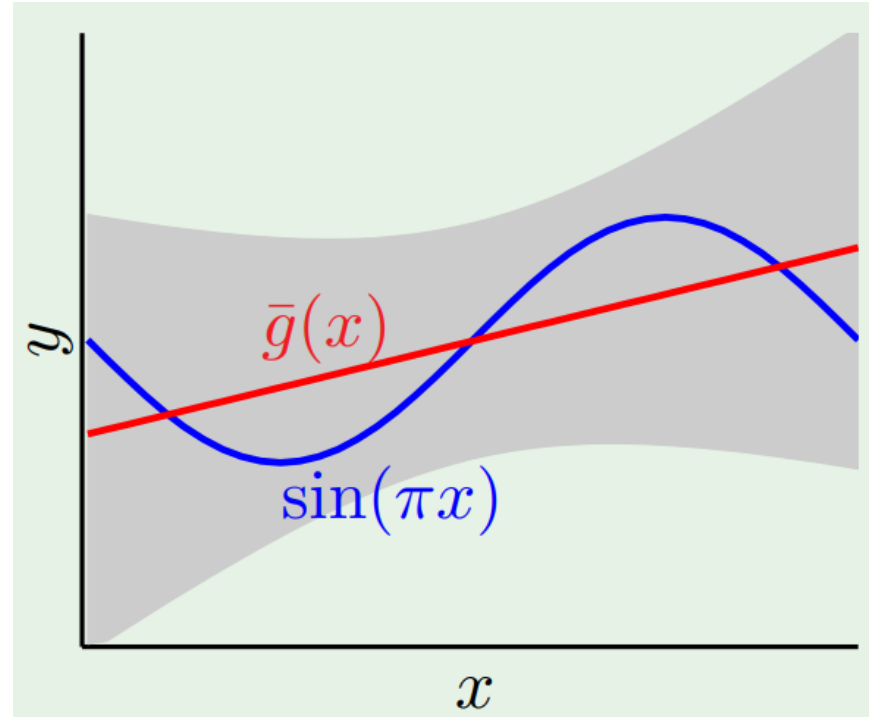
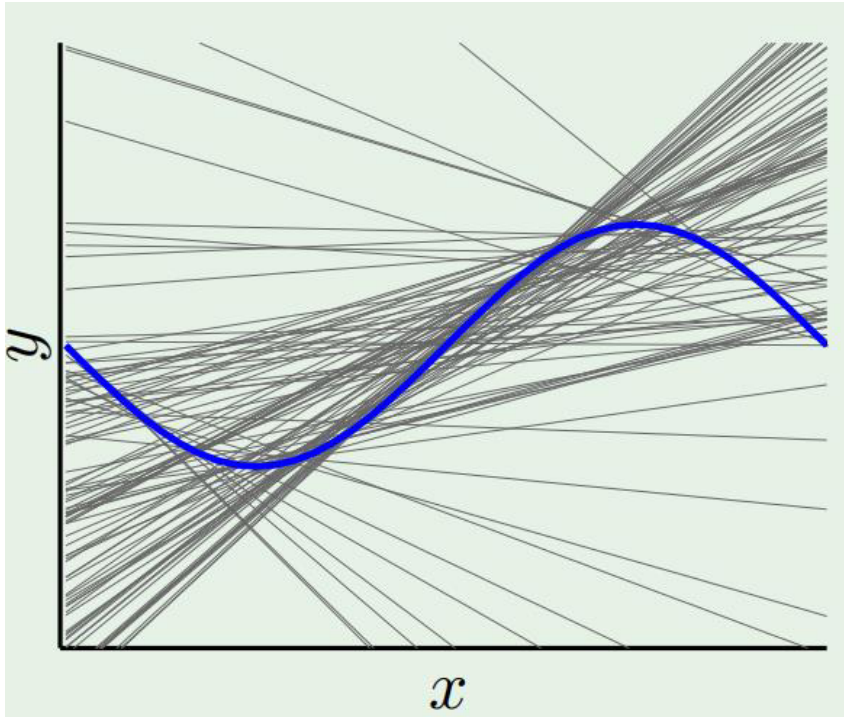
# Aprendizado - $H_0 \times H_1$



# Viés e Variância – $H_0$

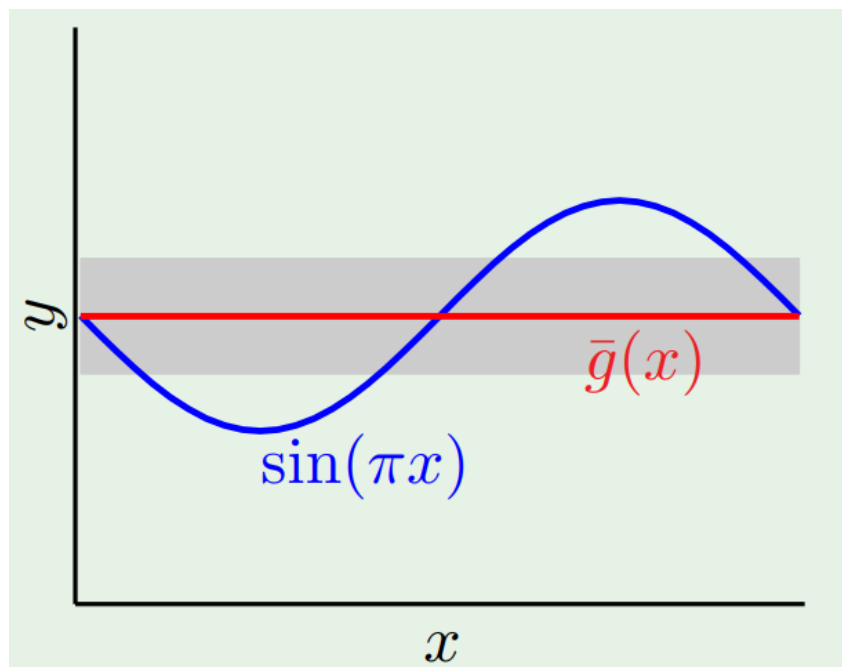


# Viés e Variância – $H_1$



$$H_0 \times H_1$$

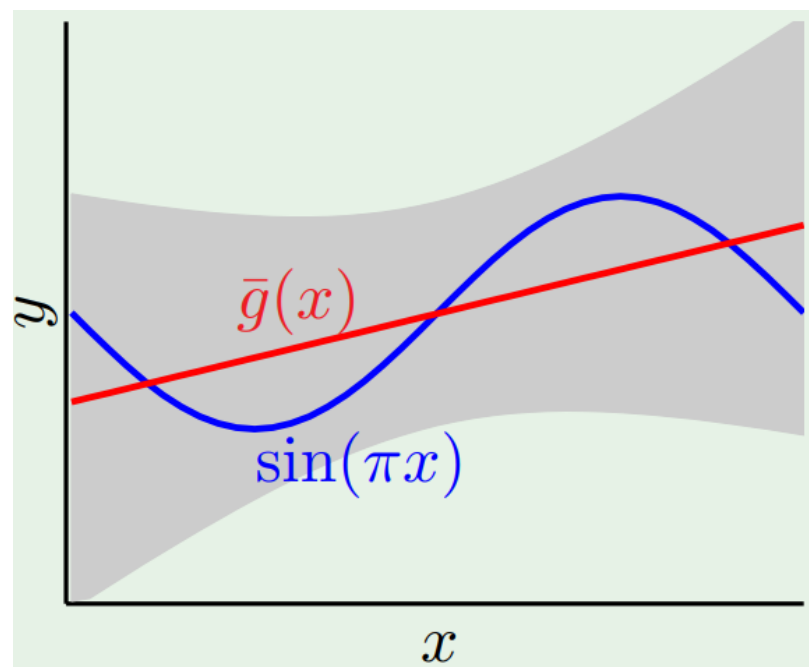
$H_0$



$$viés = 0.50$$

$$var = 0.25$$

$H_1$



$$viés = 0.21$$

$$var = 1.69$$

# Lição sobre aprendizagem

Escolha a **complexidade do modelo**

com os **recursos dos dados**,

não com a **complexidade da função alvo**

*Overfitting*

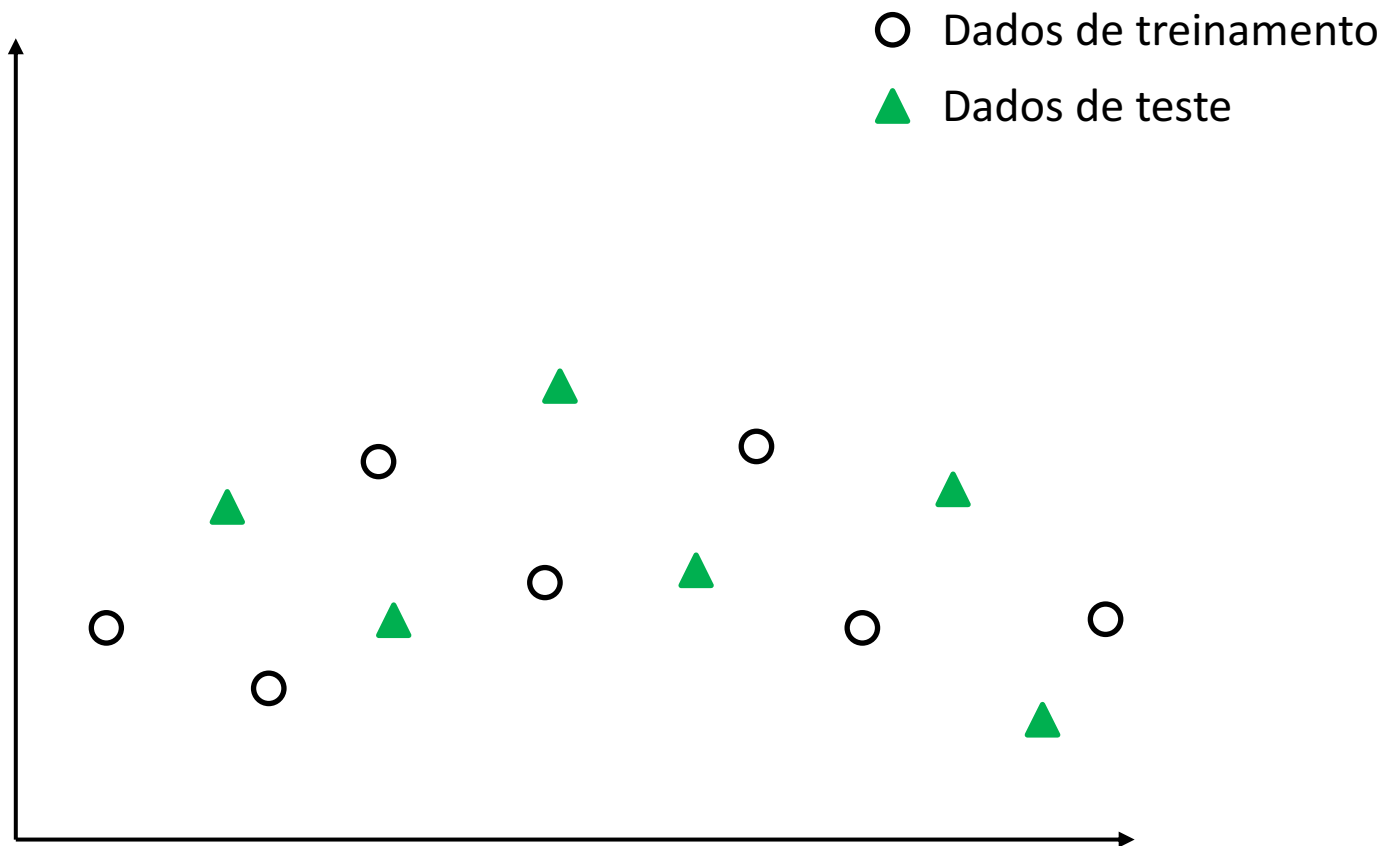
# *Overfitting*

*Overfitting*: “Aprender os dados mais que o esperado”

Ocorre quando o  $E_{in}$  é pequeno e o  $E_{out}$  é grande

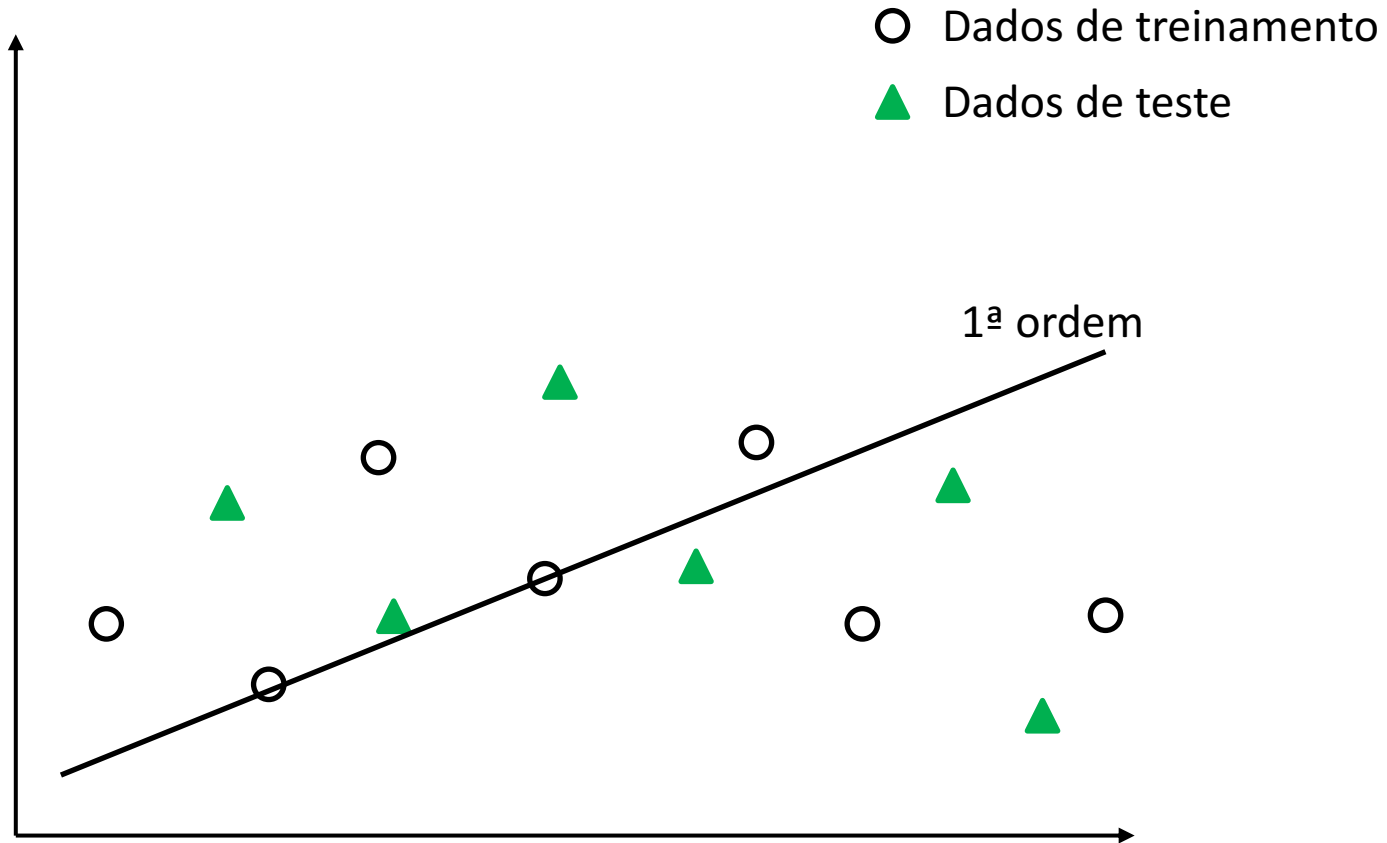
$E_{in}$  sozinho não é um bom guia para a aprendizagem

# *Overfitting*



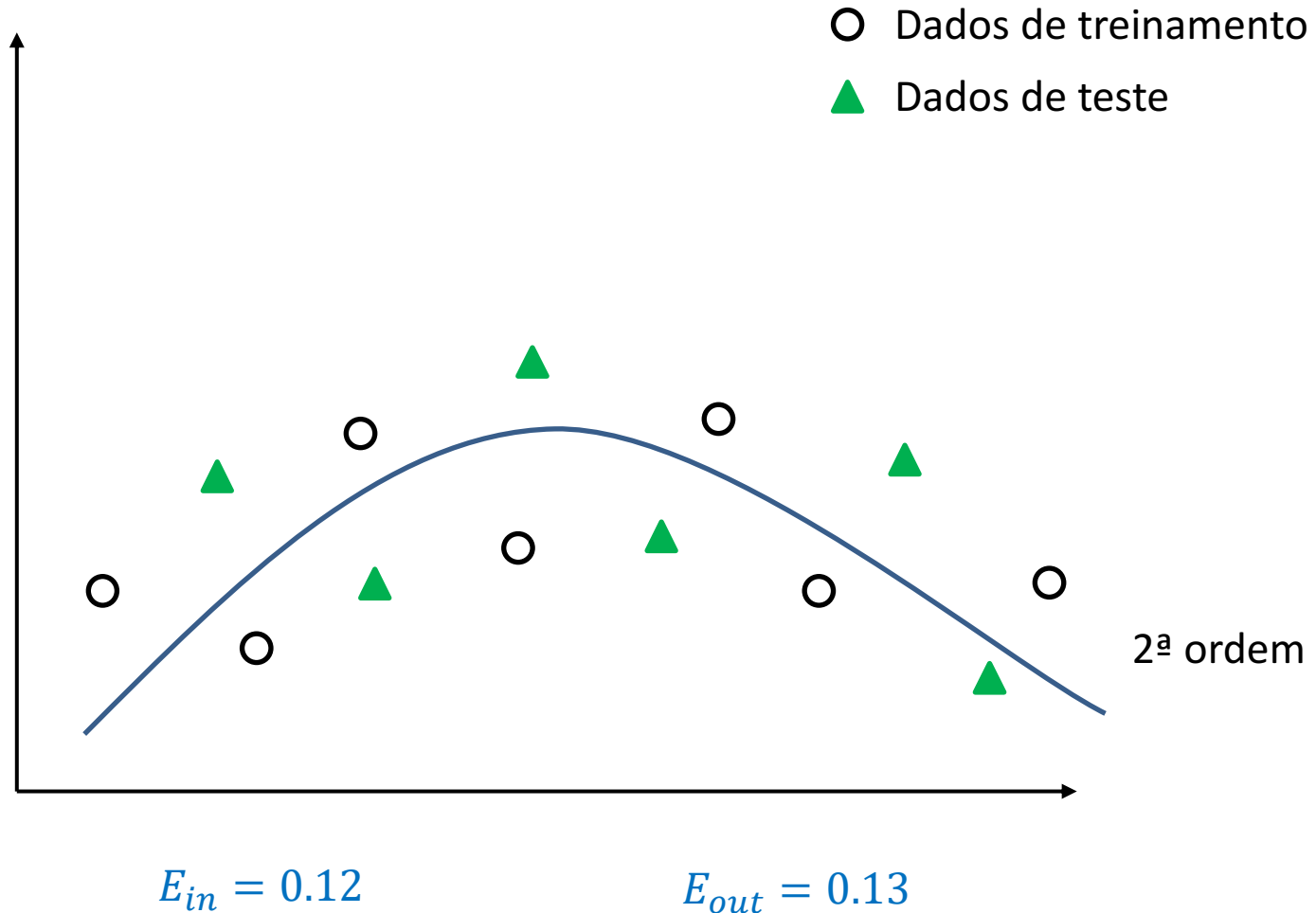


# Overfitting

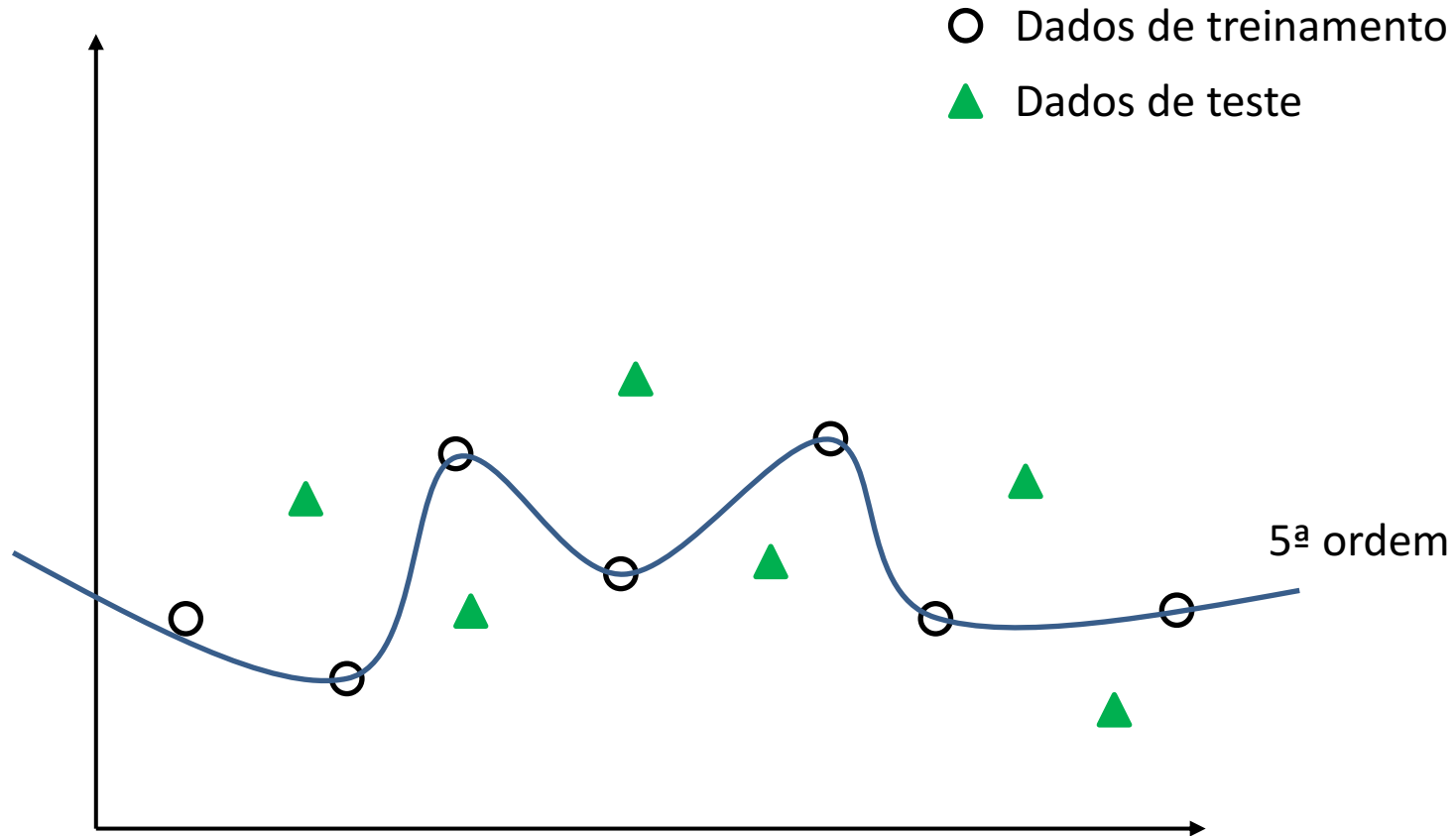


“Underfitting”

# Overfitting



# Overfitting



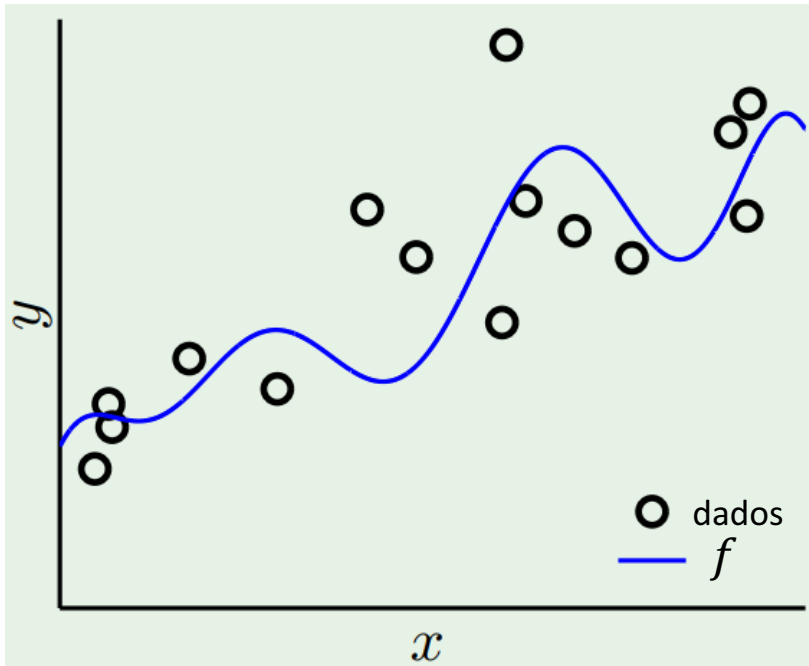
$$E_{in} = 0.1$$

$$E_{out} = 0.27$$

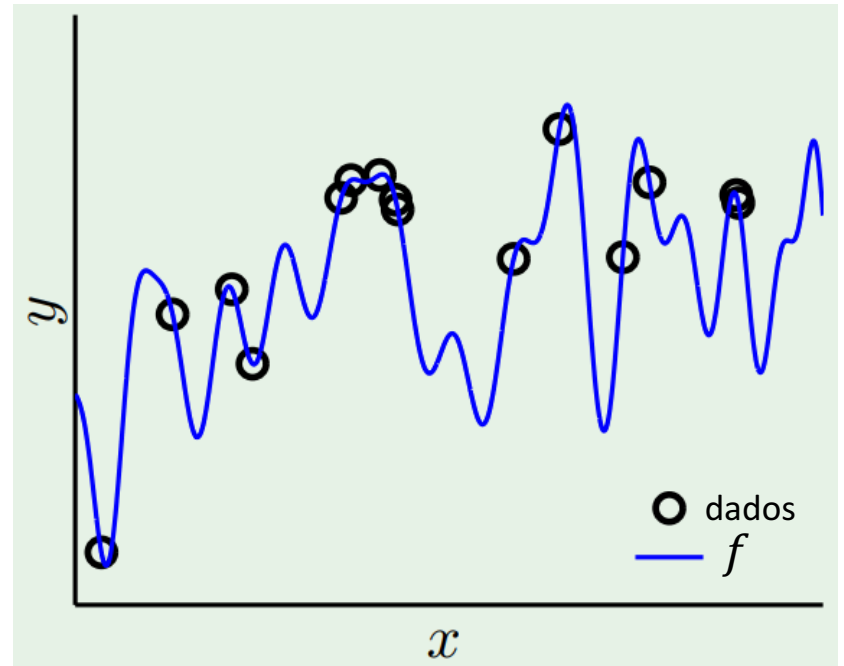
“Overfitting”

# Estudo de caso

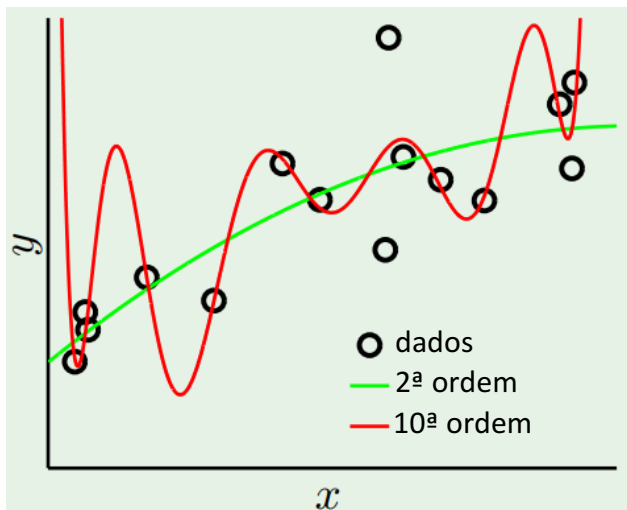
Função alvo de 10ª ordem  
+ ruído nos dados



Função alvo de 50ª ordem

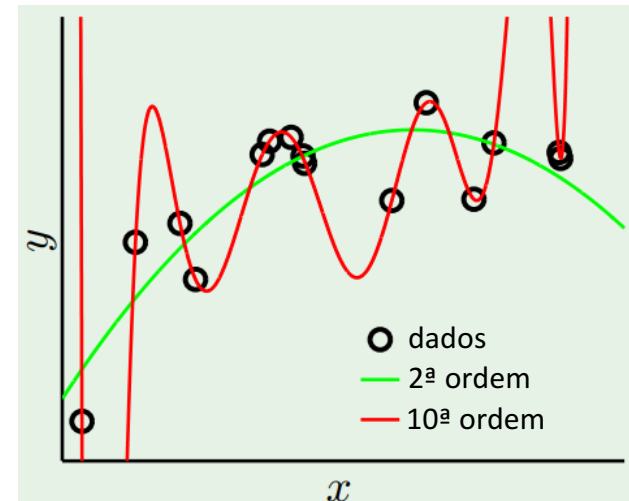


# Ajustes para os alvos



Com ruído – Alvo 10ª ordem

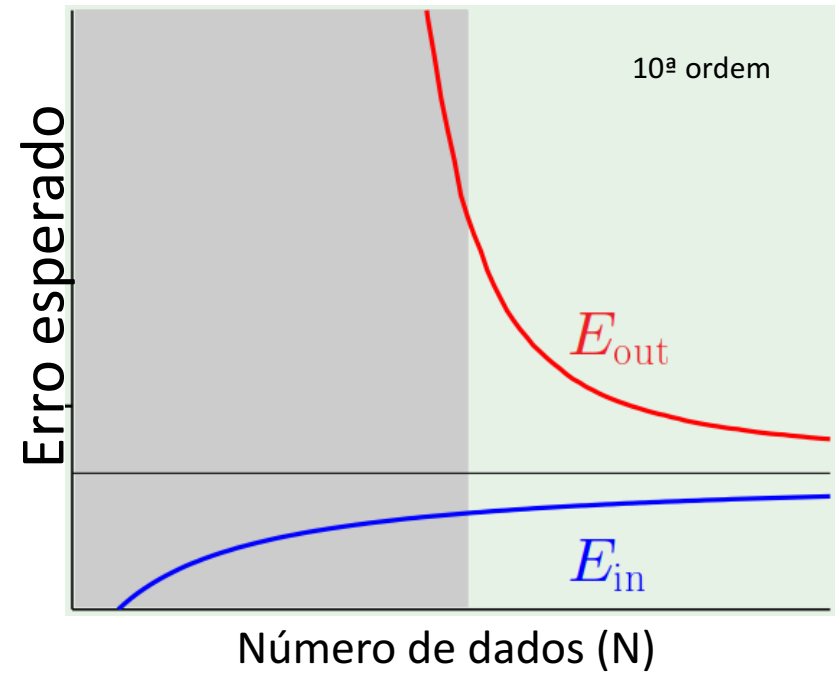
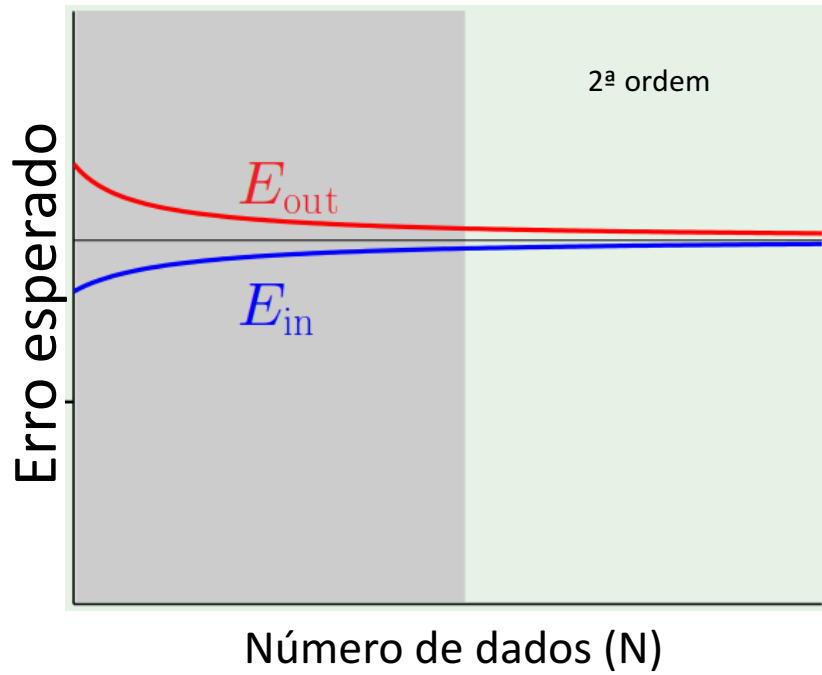
|           | 2ª ordem | 10ª ordem |
|-----------|----------|-----------|
| $E_{in}$  | 0,05     | 0,0034    |
| $E_{out}$ | 0,127    | 9,00      |



Sem ruído – Alvo 50ª ordem

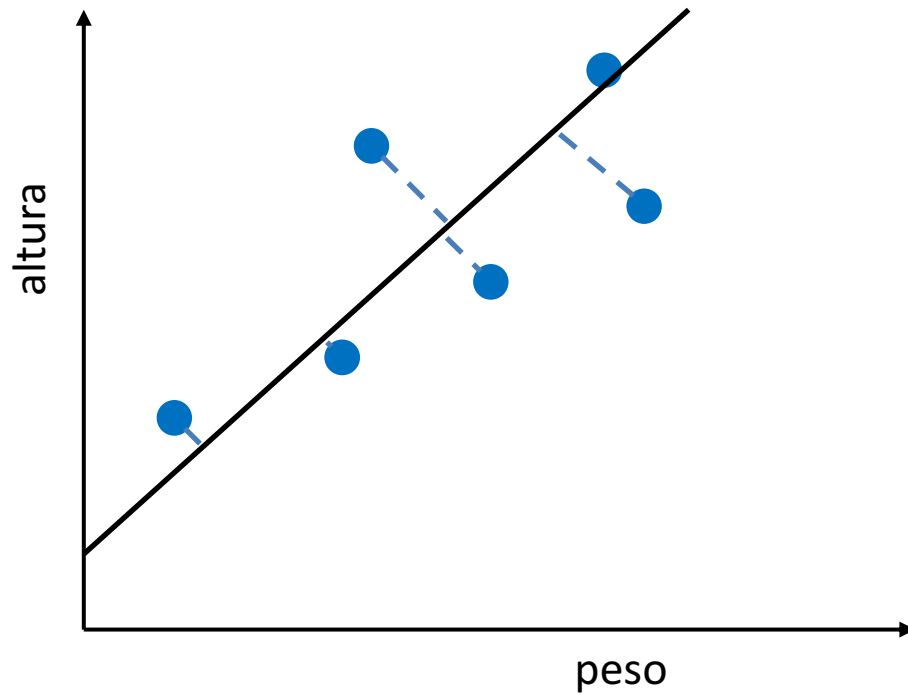
|           | 2ª ordem | 10ª ordem |
|-----------|----------|-----------|
| $E_{in}$  | 0,029    | $10^{-5}$ |
| $E_{out}$ | 0,120    | 7680      |

# Curva de Aprendizizado

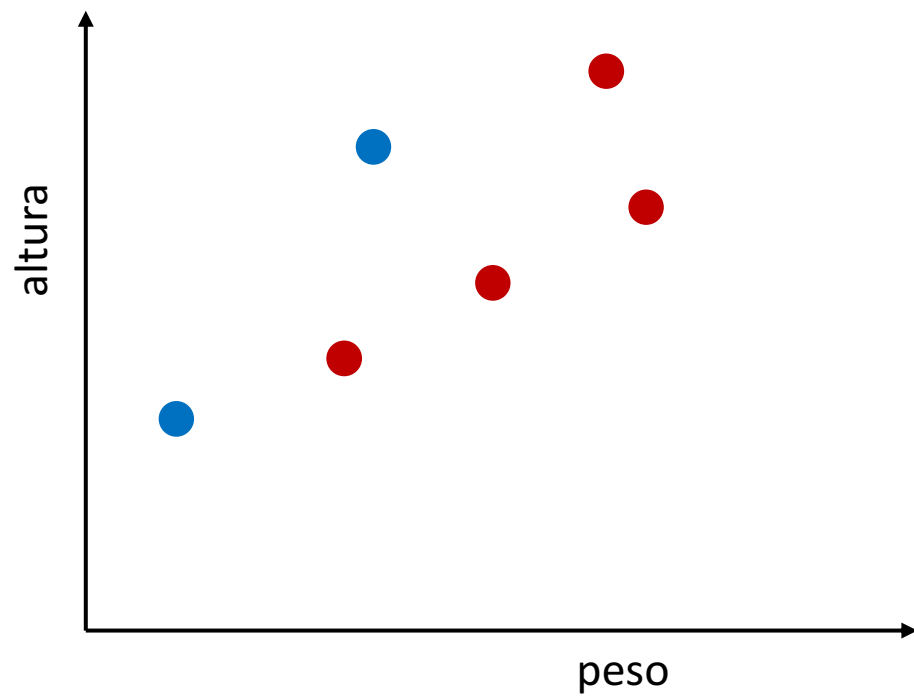


*Regularização*

$$\begin{aligned} altura &= 0,9 + 0,75 \cdot peso \\ g(x) &= w_0 + w_1 \cdot peso \end{aligned}$$



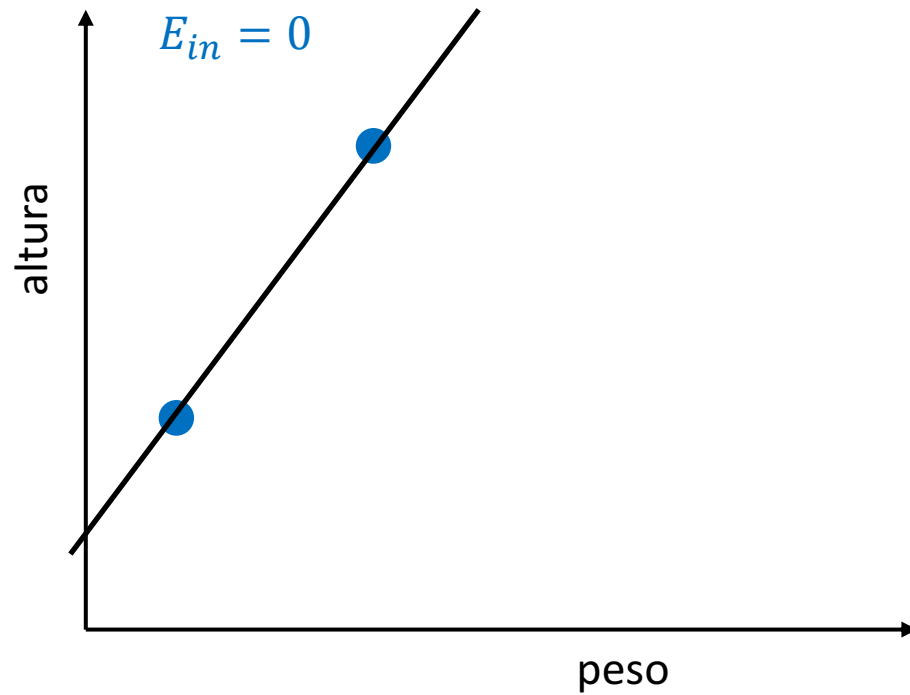




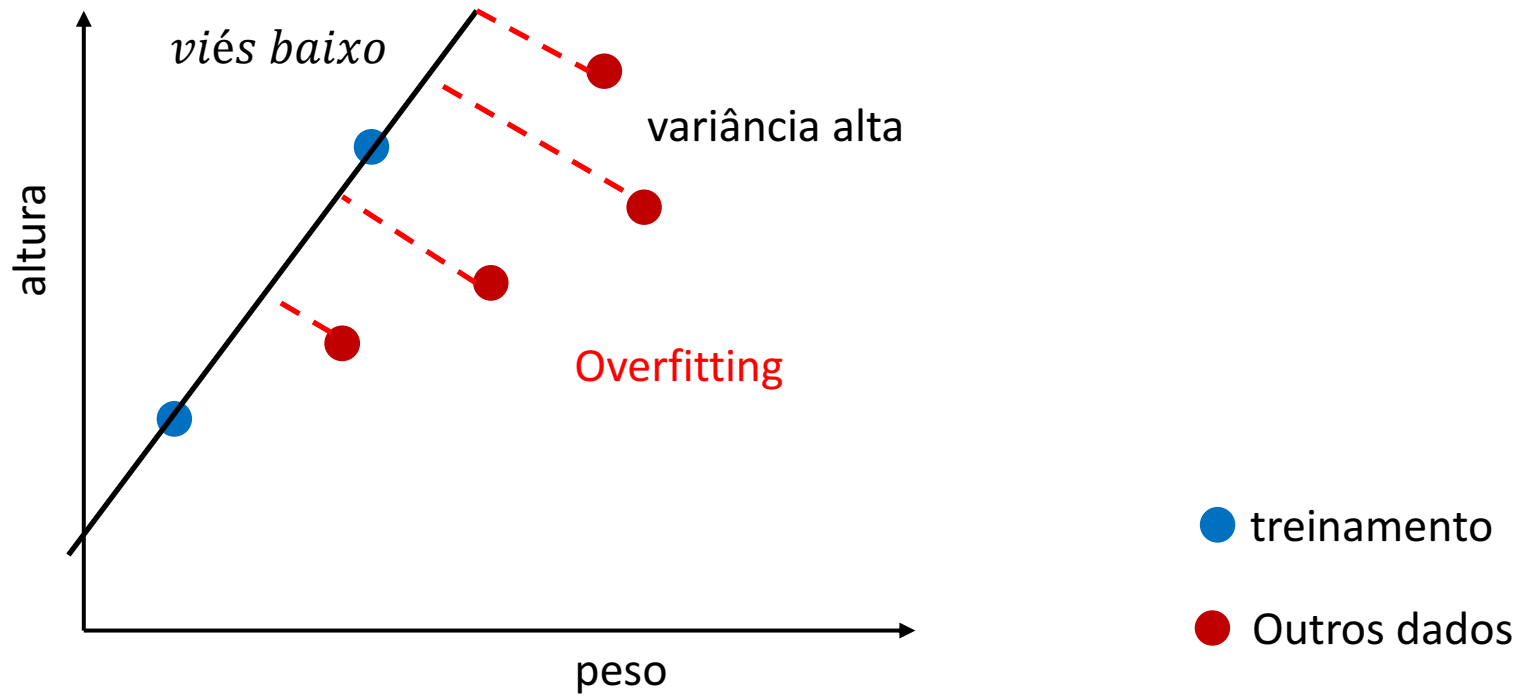
● treinamento

● Outros dados

$$altura = 0,4 + 1,3 \cdot peso$$



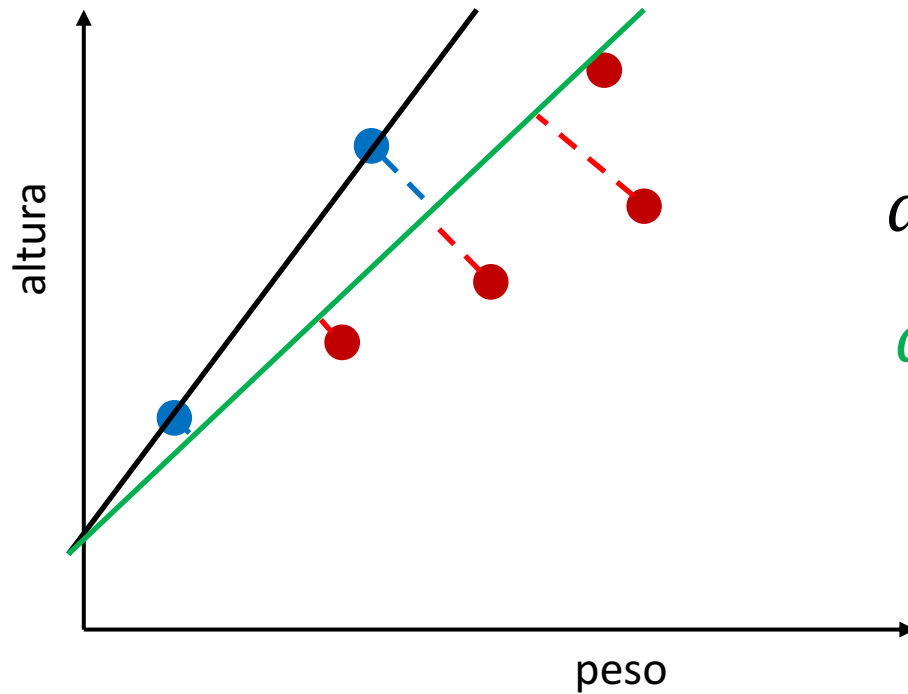
$$altura = 0,4 + 1,3 \cdot peso$$



# Regularização

Ideia:

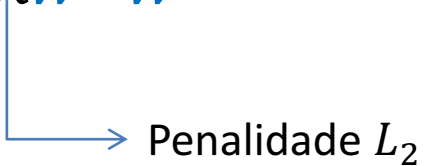
- Aumentar um pouco o viés
- Diminuir a Variância



$$altura = 0,4 + 1,3 \cdot peso$$

$$altura = 0,4 + 0,9 \cdot peso$$

# Regularização - Penalidade $L_2$

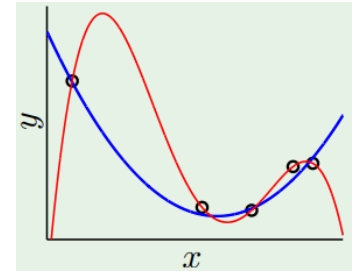
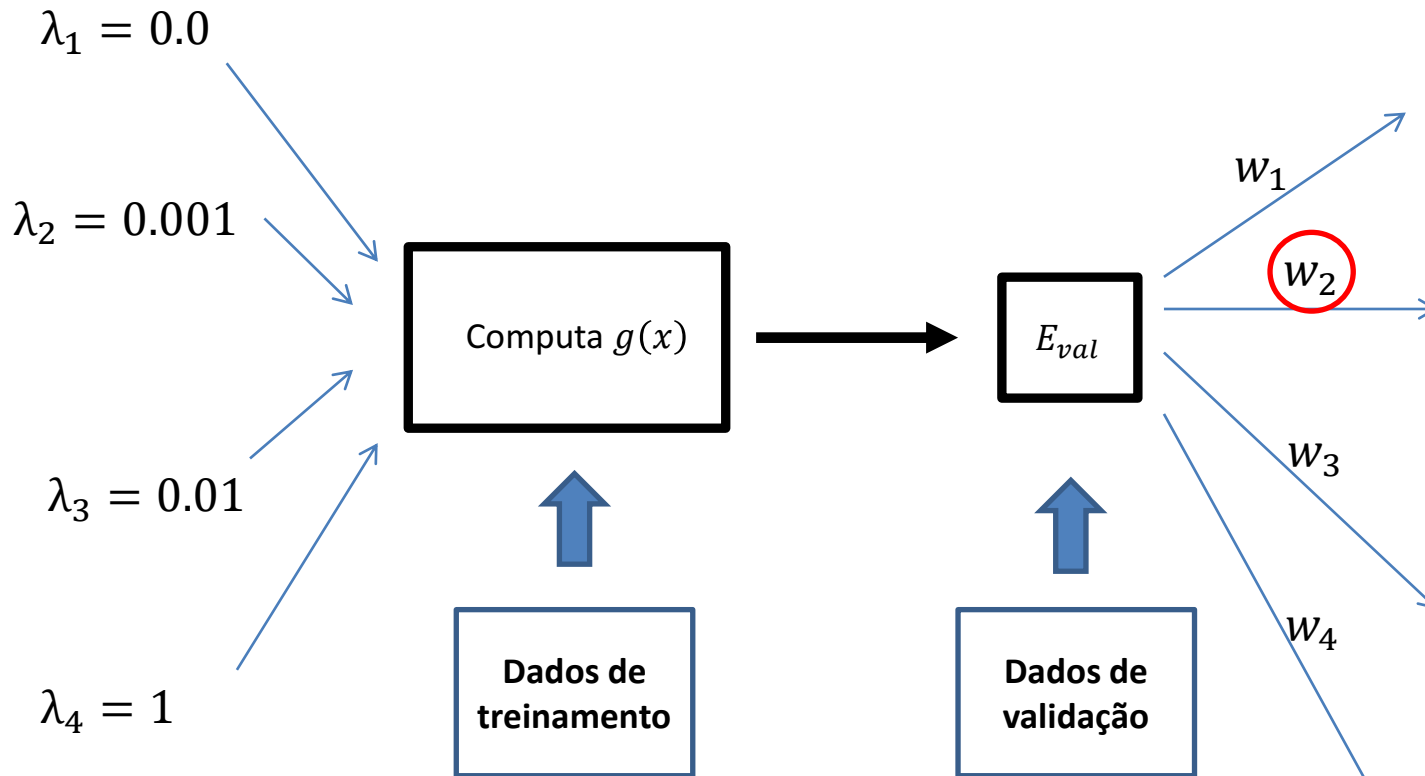
$$E_{aug}(\mathbf{w}) = E_{in}(\mathbf{w}) + \lambda \mathbf{w}^T \mathbf{w}$$


Penalidade  $L_2$

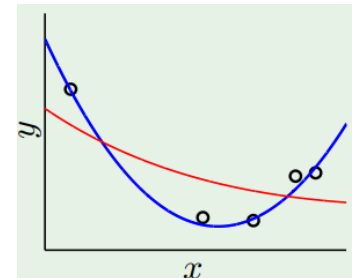
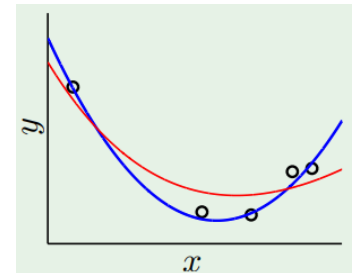
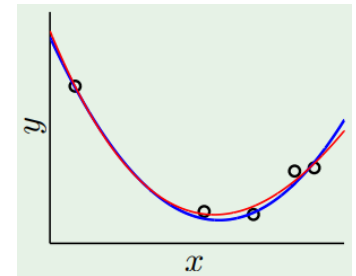
Gradiente Descendente:

$$\begin{array}{l} \text{Total} \\ \text{Derivada} \end{array} = \nabla E_{in}(\mathbf{w}(t)) + 2 \lambda \mathbf{w}(t)$$

# Heurística Weight Decay



Overfitting



Underfitting