

# Machine Learning Aplicado à Cartografia Geológica

## Como a Máquina Aprende?

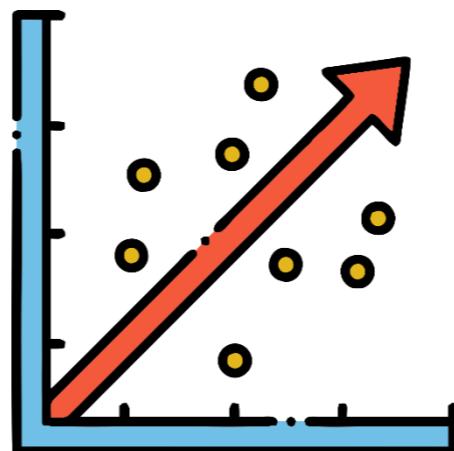


Iago Costa

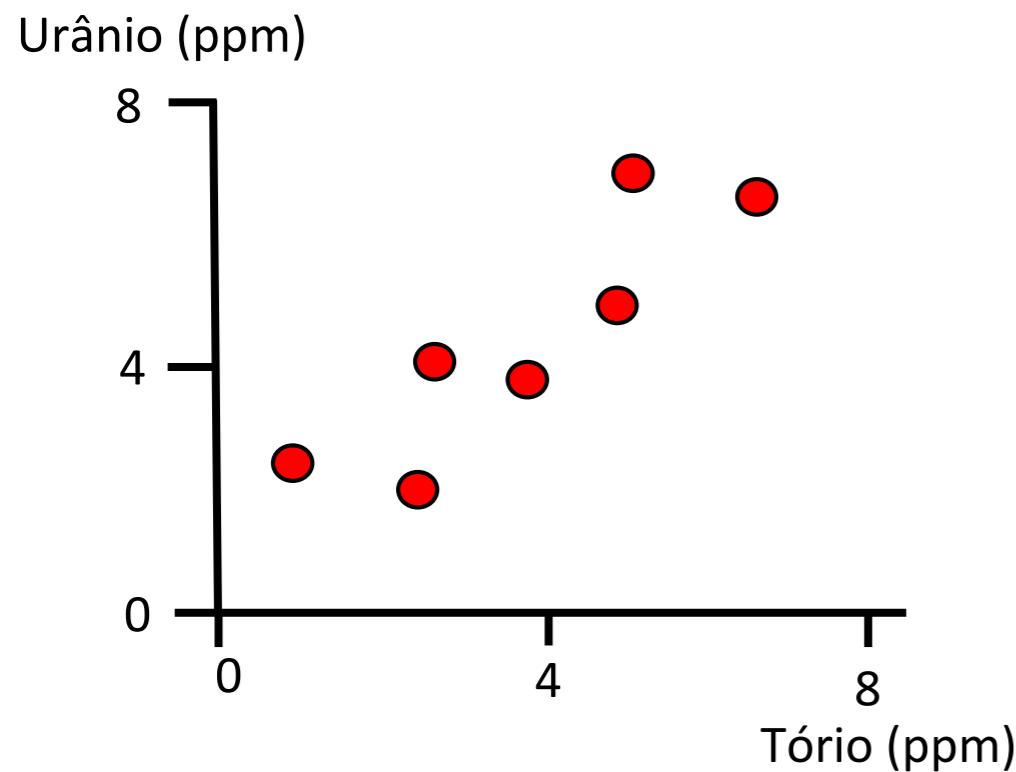
IX SimBGF

4 de outubro de 2022

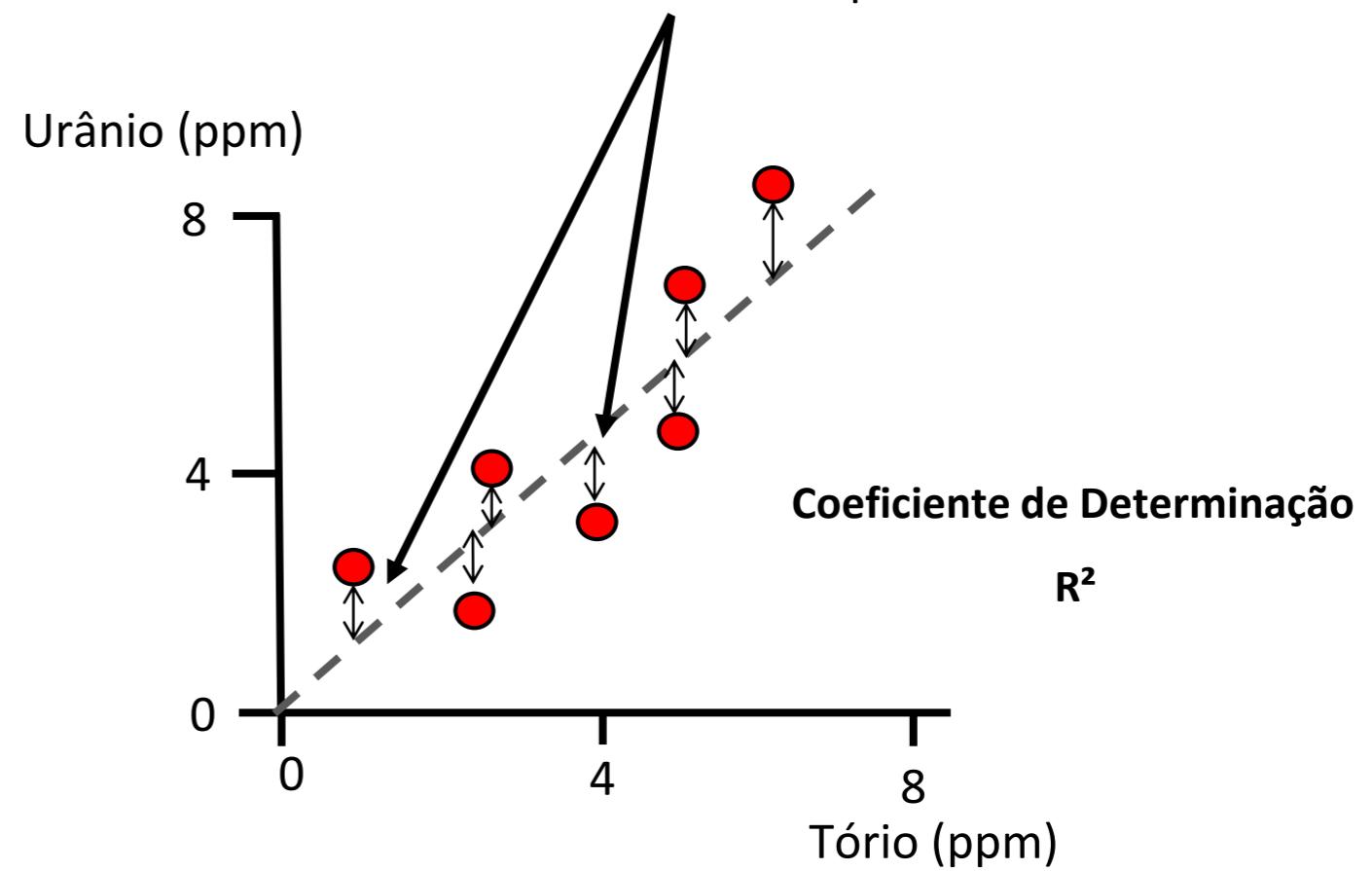
# Algoritmos Lineares



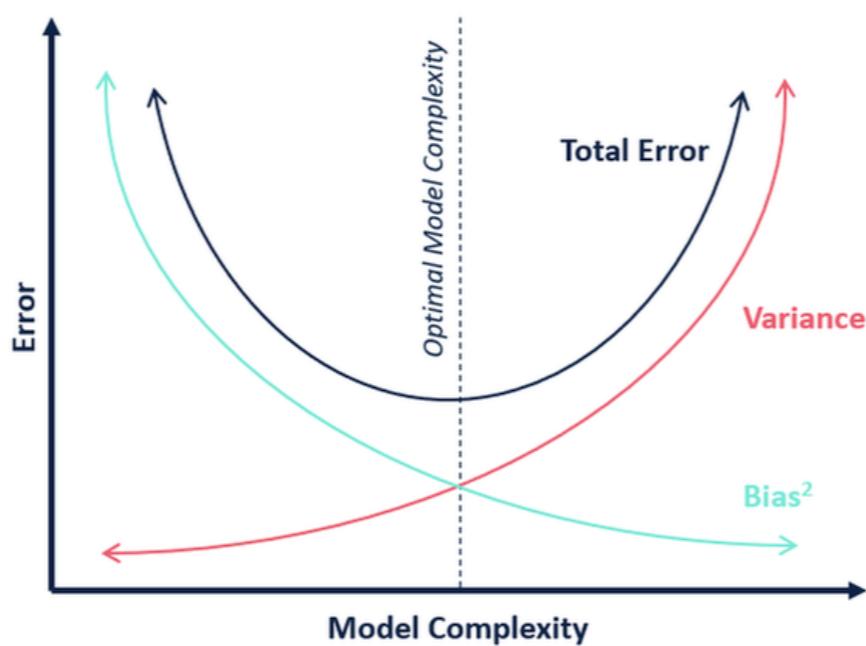
# Regressão Linear



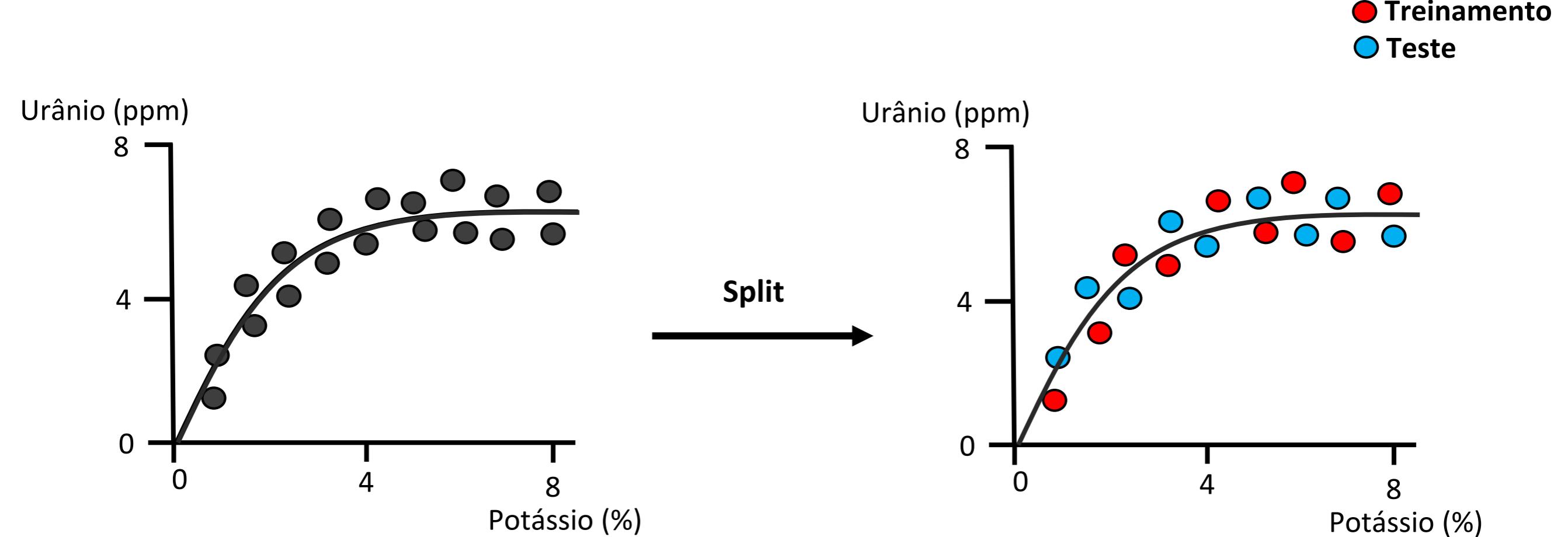
**Método dos mínimos Quadrados**  
minimiza a soma dos quadrados dos Erros!



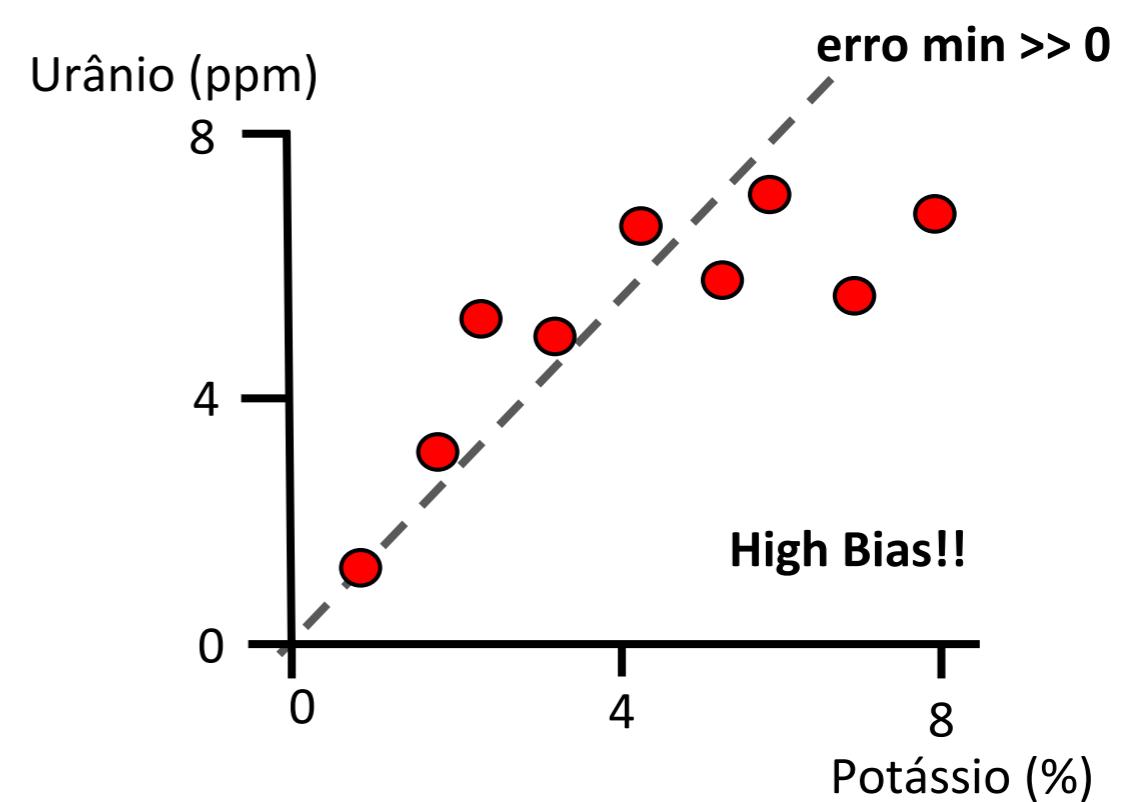
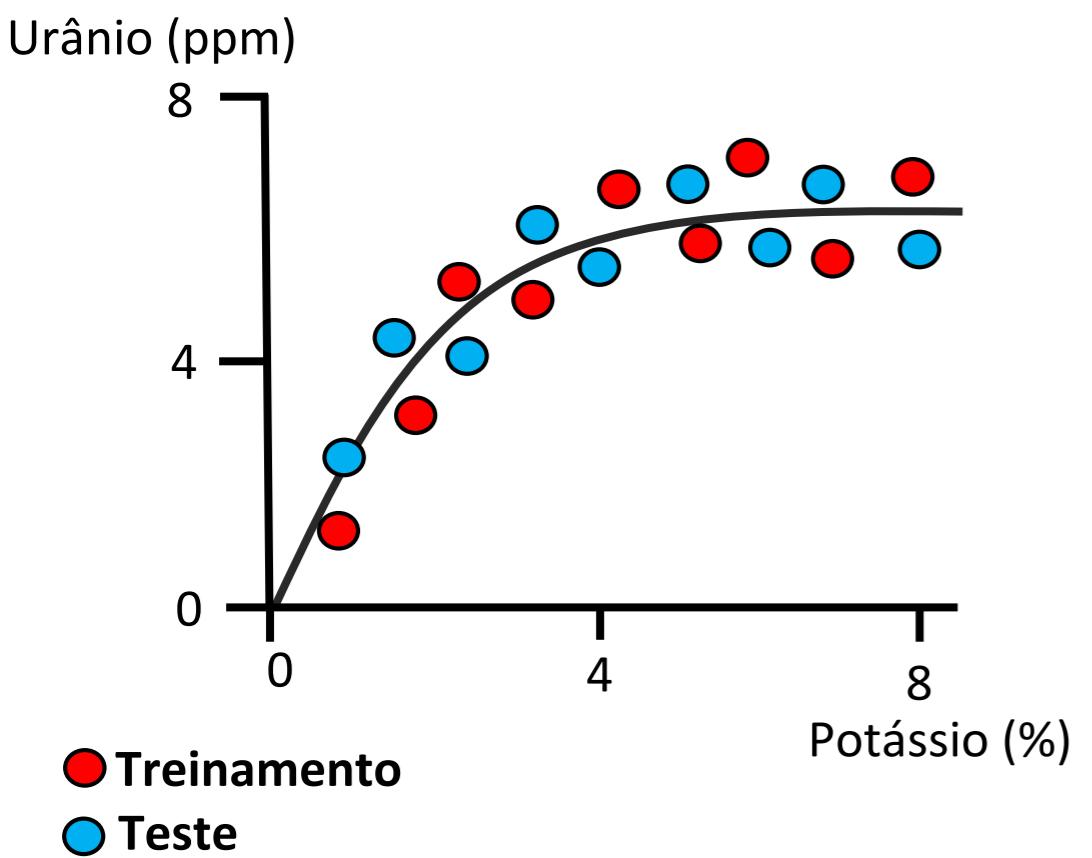
# Bias vs Variance



# Bias vs Variance

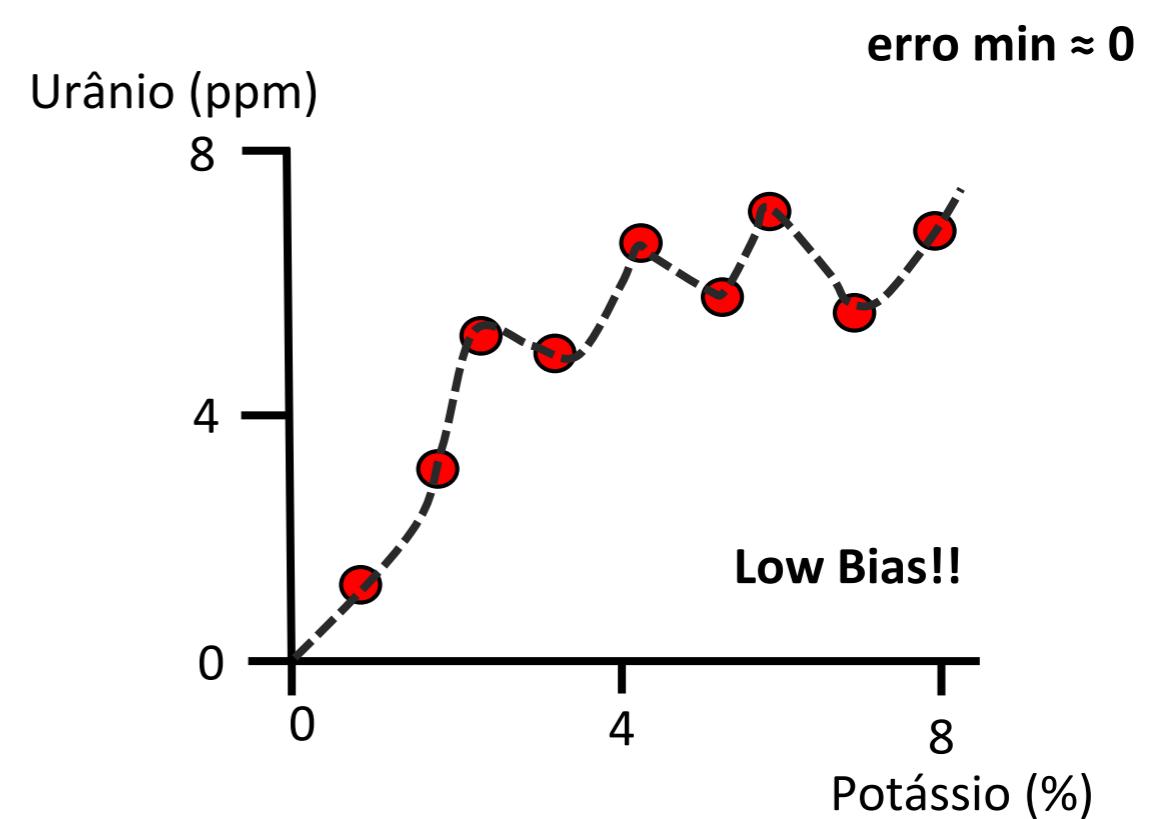
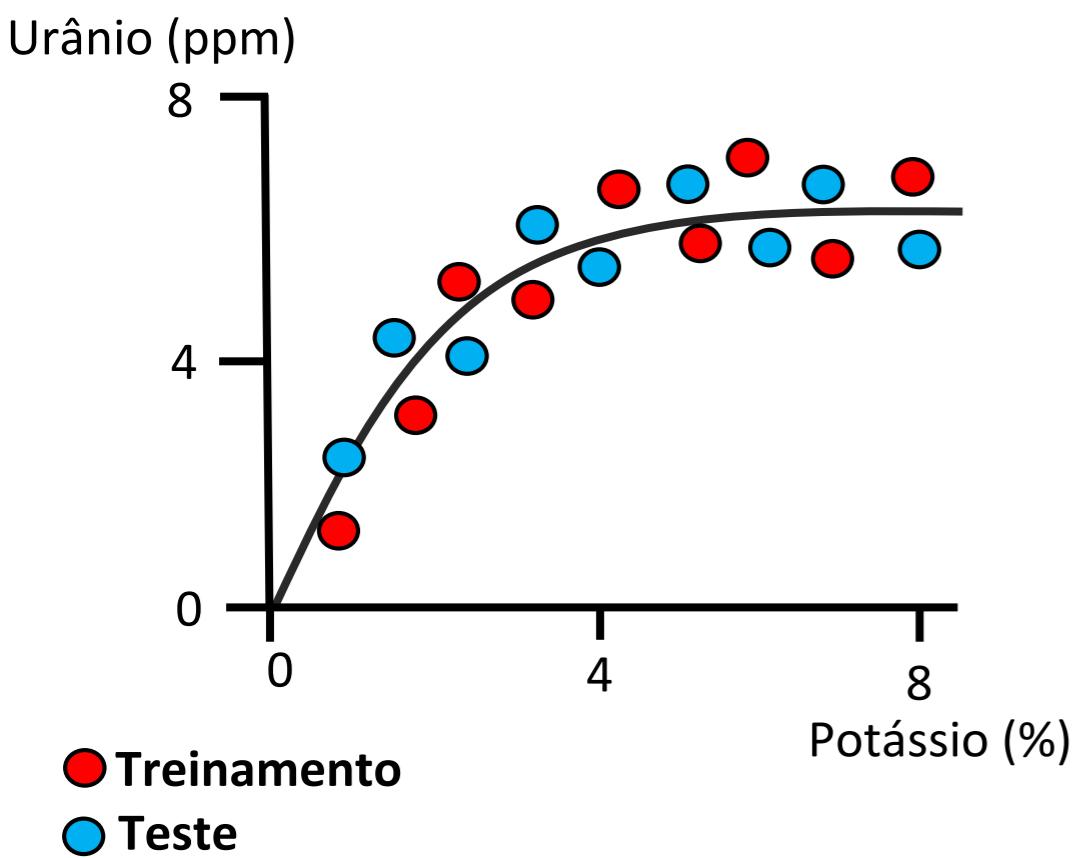


# Bias vs Variance



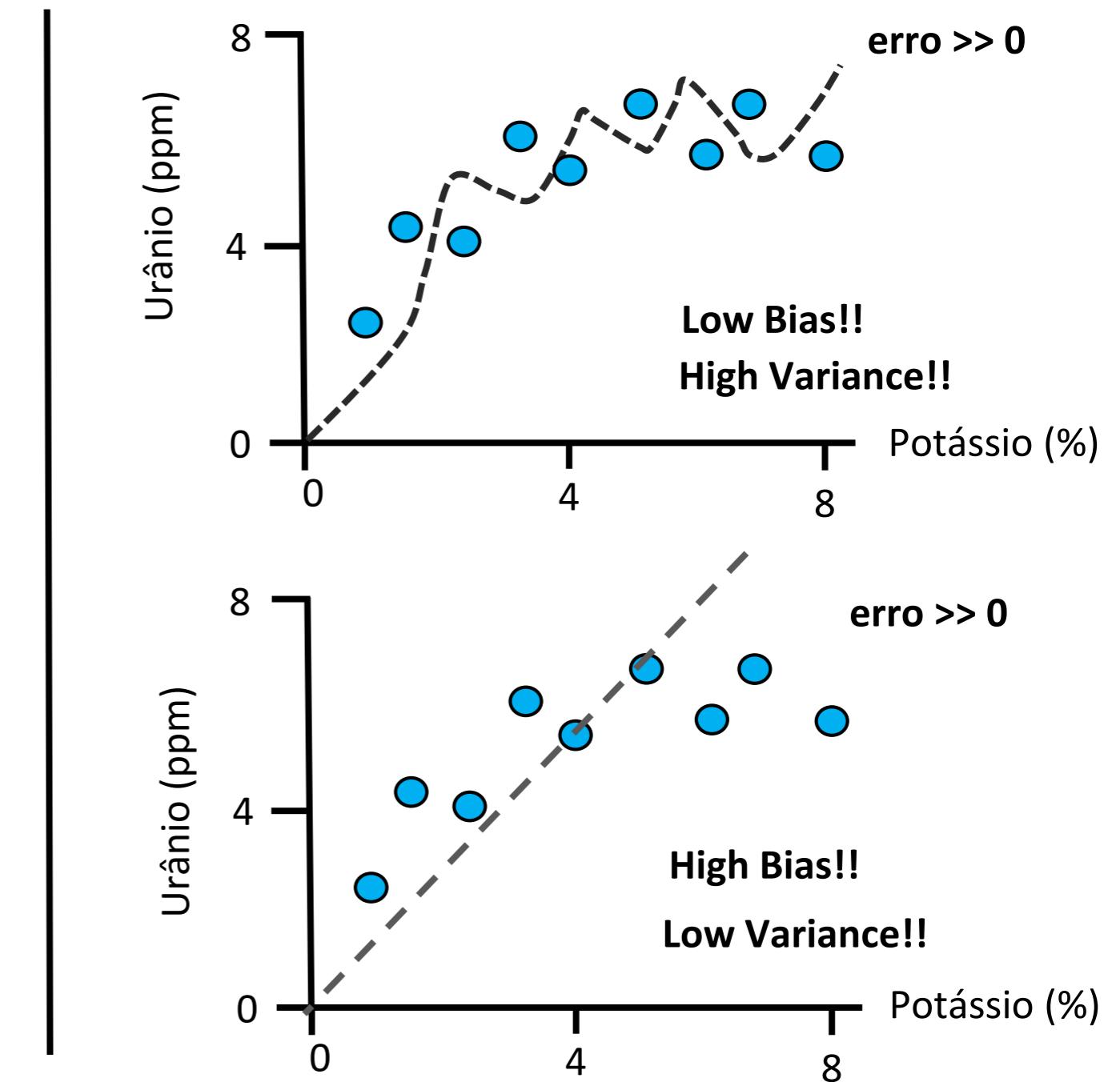
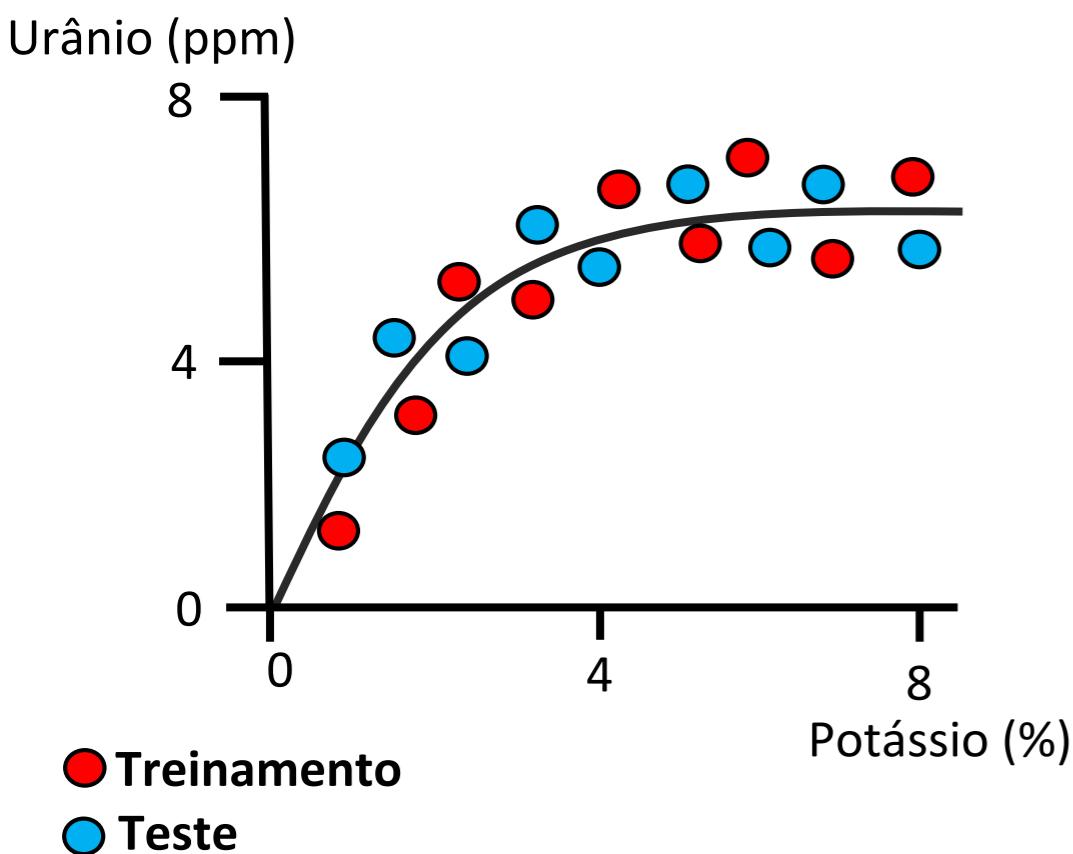
Bias é a diferença entre a previsão  
do modelo e o dado real!

# Bias vs Variance



Bias é a diferença entre a previsão  
do modelo e o dado real!

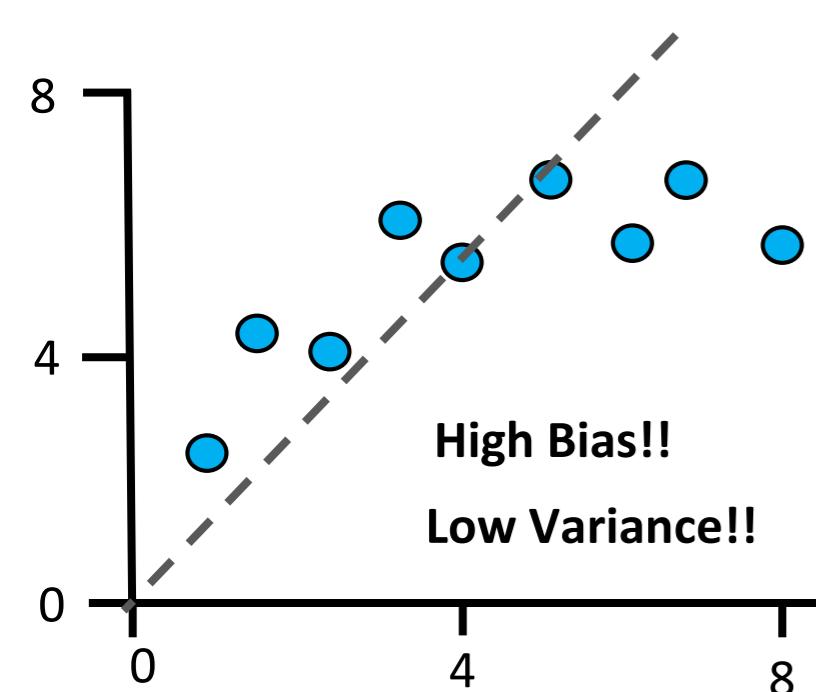
# Bias vs Variance



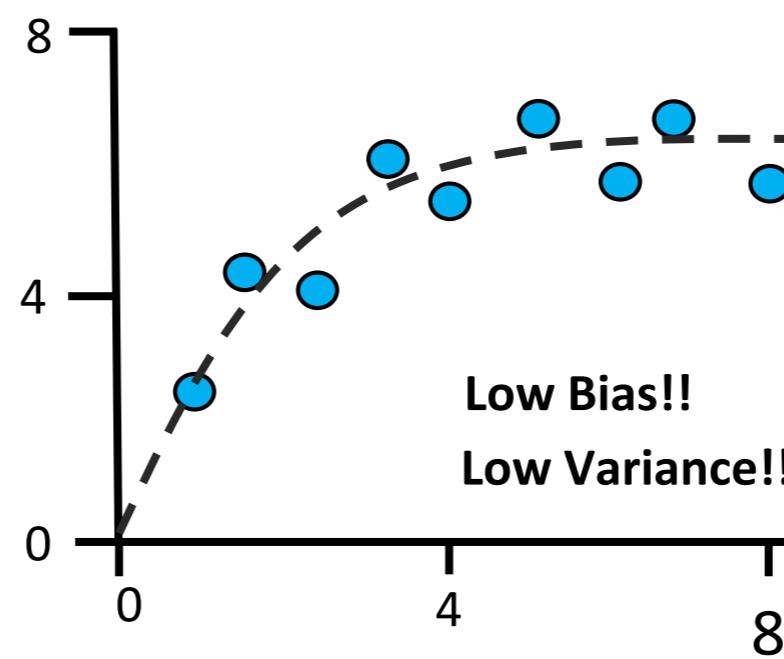
**Variance** é a diferença de ajuste entre  
diferentes conjuntos de dados

# Bias vs Variance

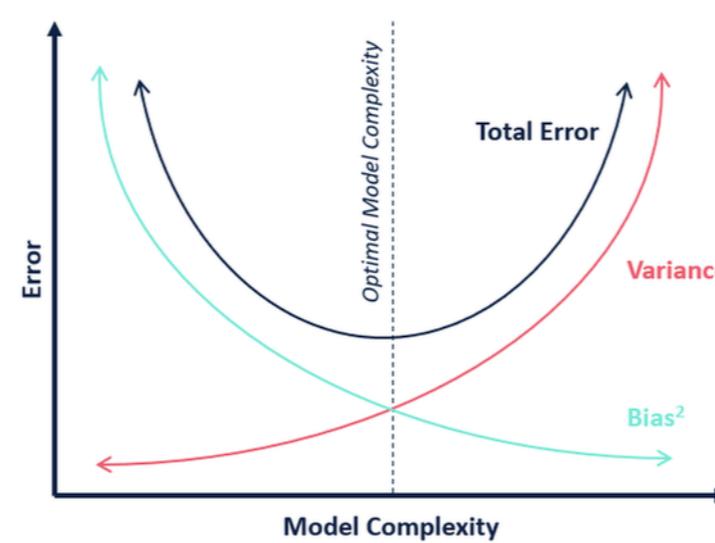
*Underfitting*



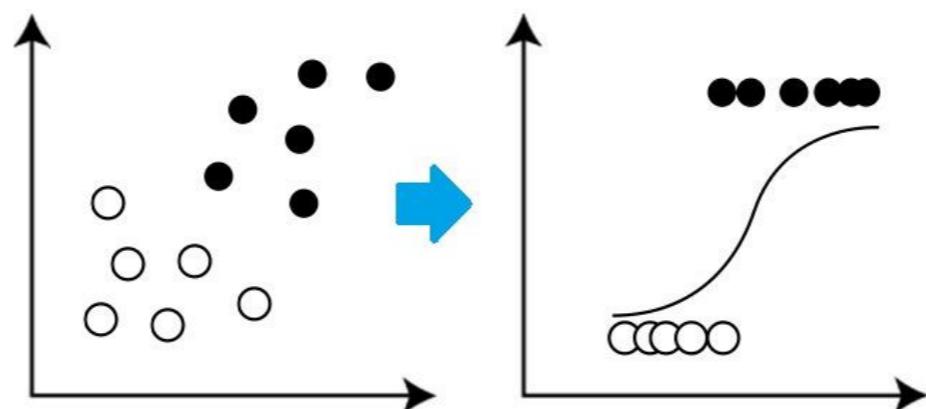
*Overfitting*



**Regularização**



# Regressão Logística (Logistic Regression)

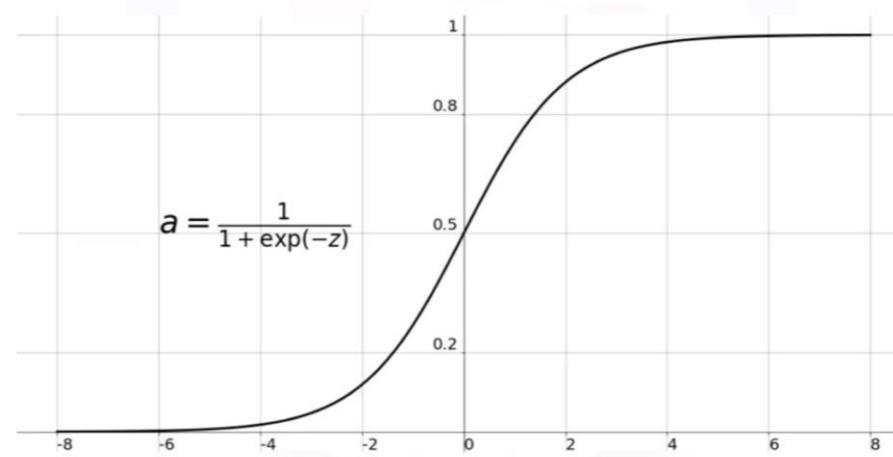


# Regressão Logística (Logistic Regression)

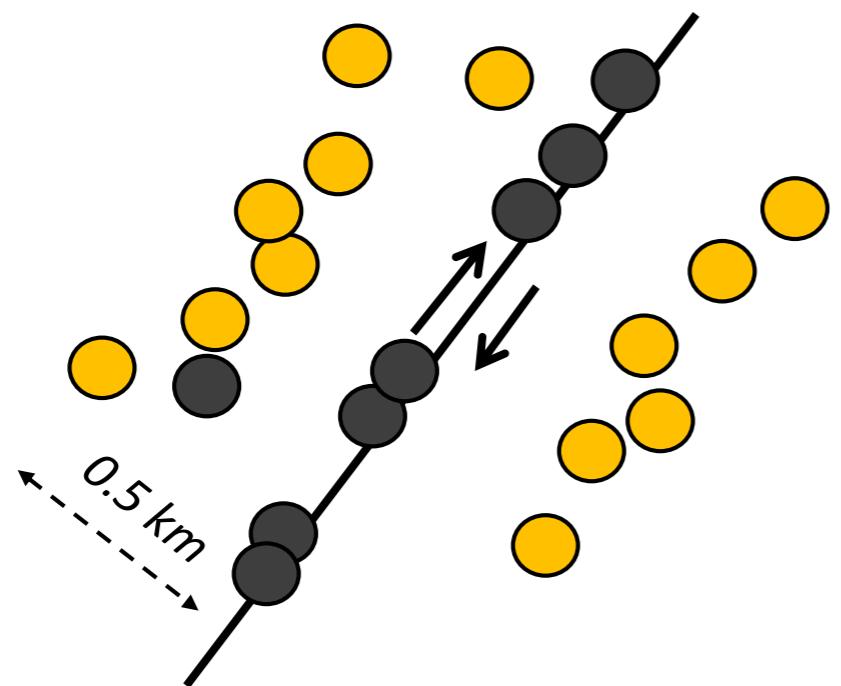
## Principais diferenças com a regressão linear

- A regressão logística faz a previsão se um evento é verdadeiro ou falso (**True or False**), enquanto a regressão linear faz previsões de valores contínuos (ex: concentração de urânio)
- A regressão logística ajusta uma função sigmoidal aos dados ao invés de uma reta.

### Sigmoid Function

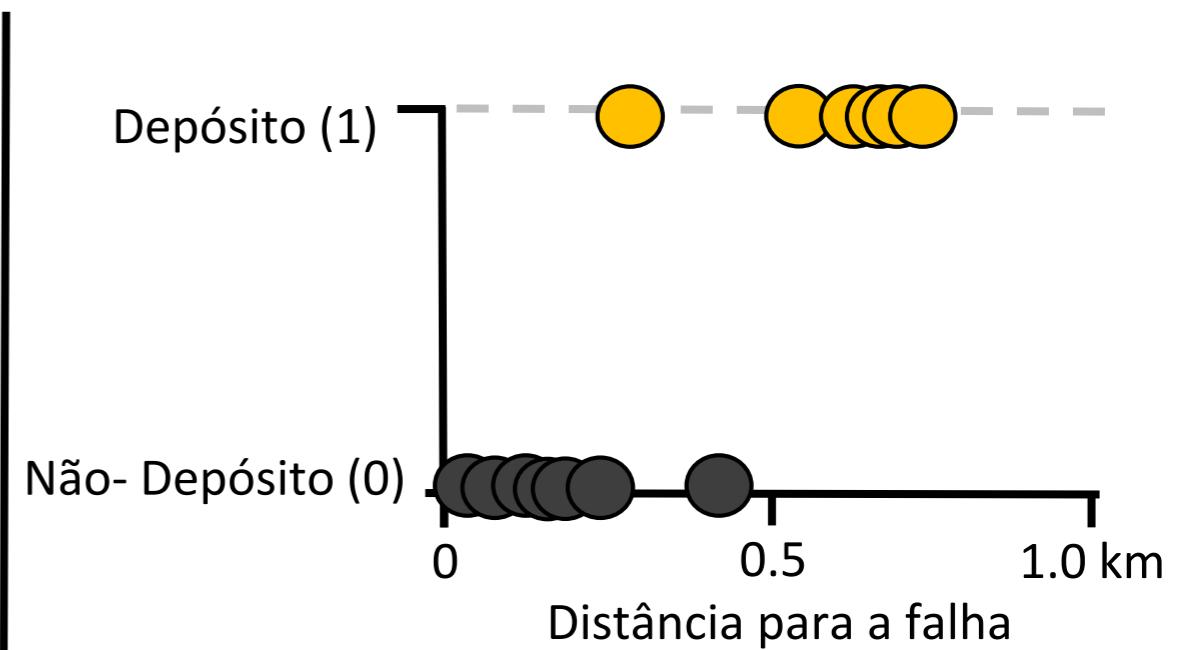


# Regressão Logística (Logistic Regression)



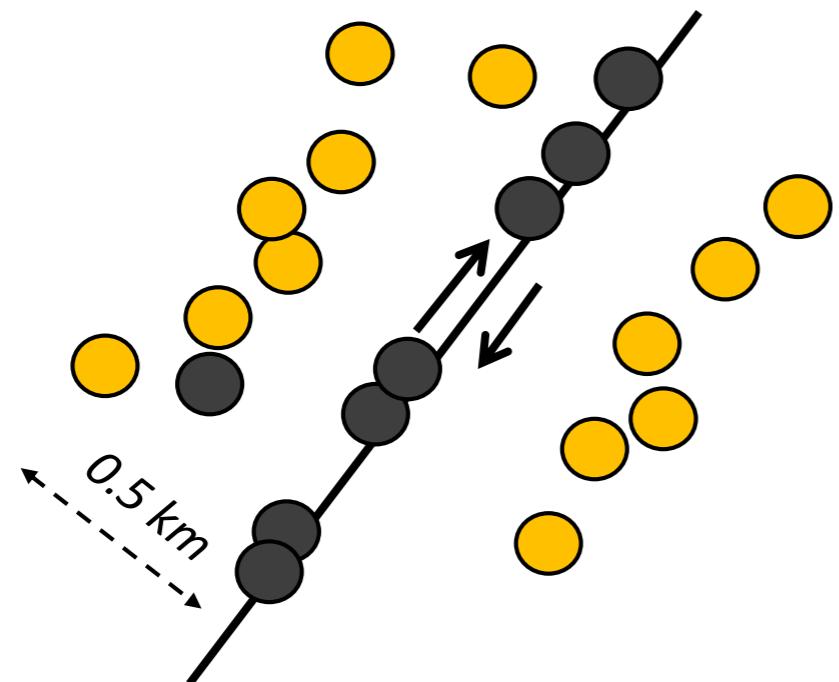
Depósito

Não – Depósito (estéril)



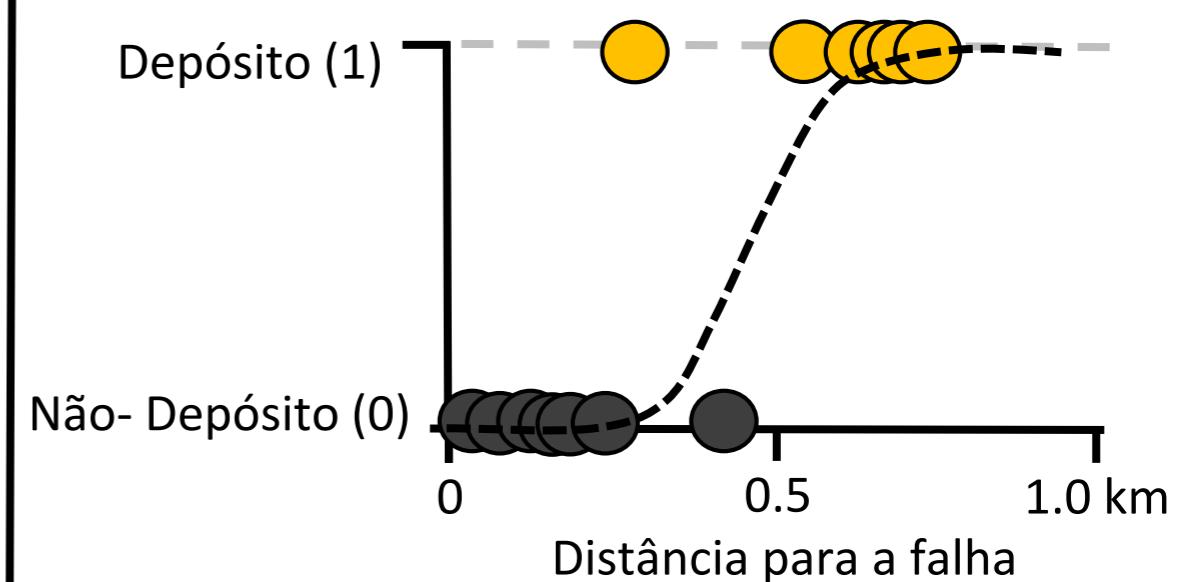
# Regressão Logística (Logistic Regression)

## Exemplo



Depósito

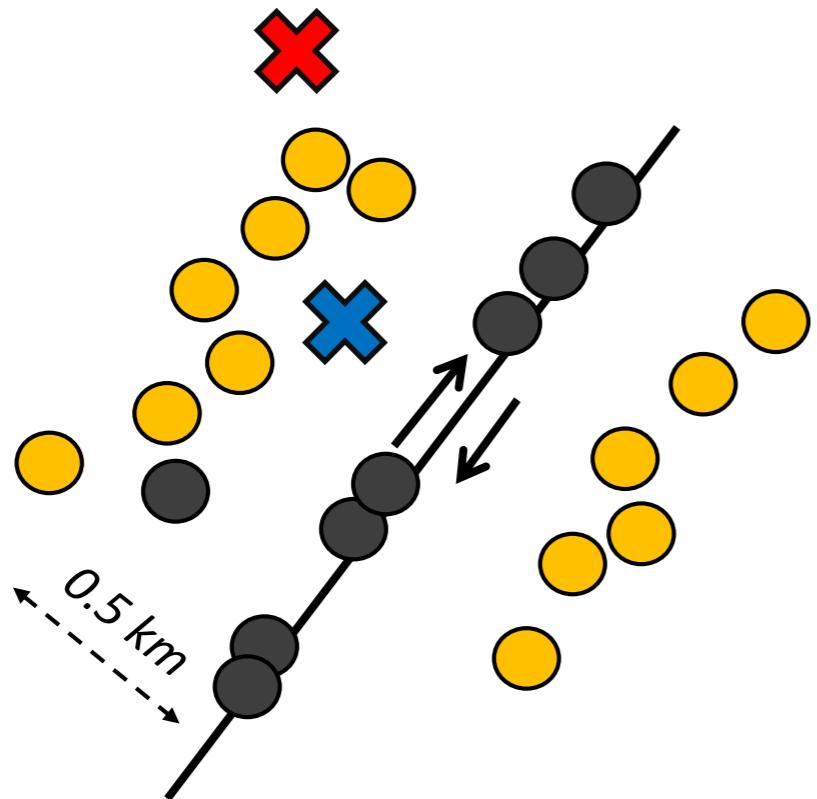
Não – Depósito (estéril)



Nota: A Regressão logística não utiliza o  $R^2$  para o ajuste da função, ela utiliza o método **Maximum likelihood!!**

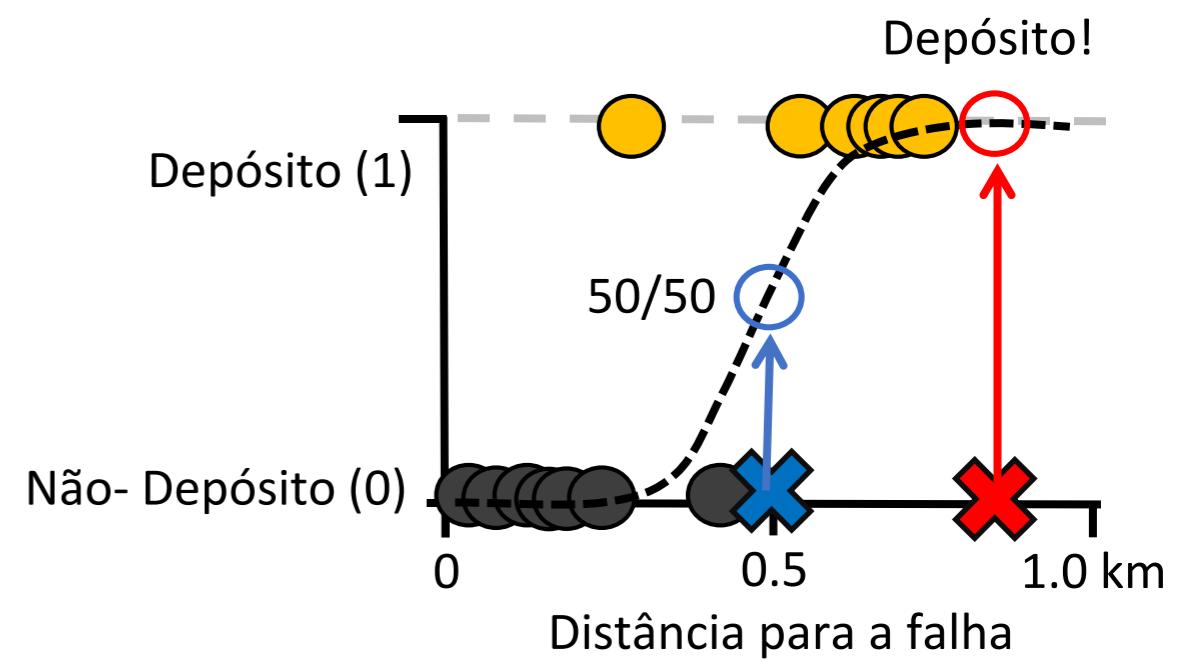
# Regressão Logística (Logistic Regression)

## Exemplo



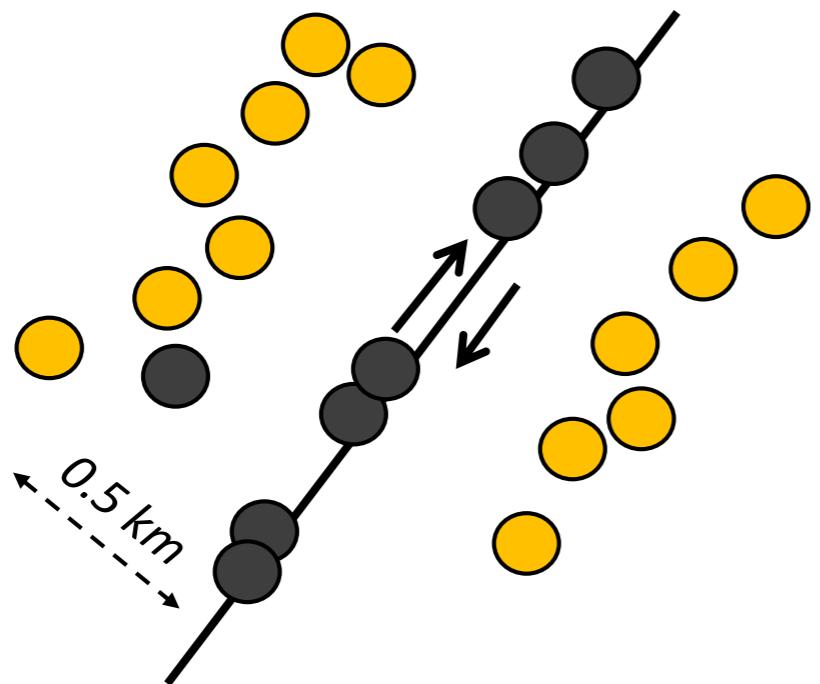
Depósito

Não – Depósito (estéril)



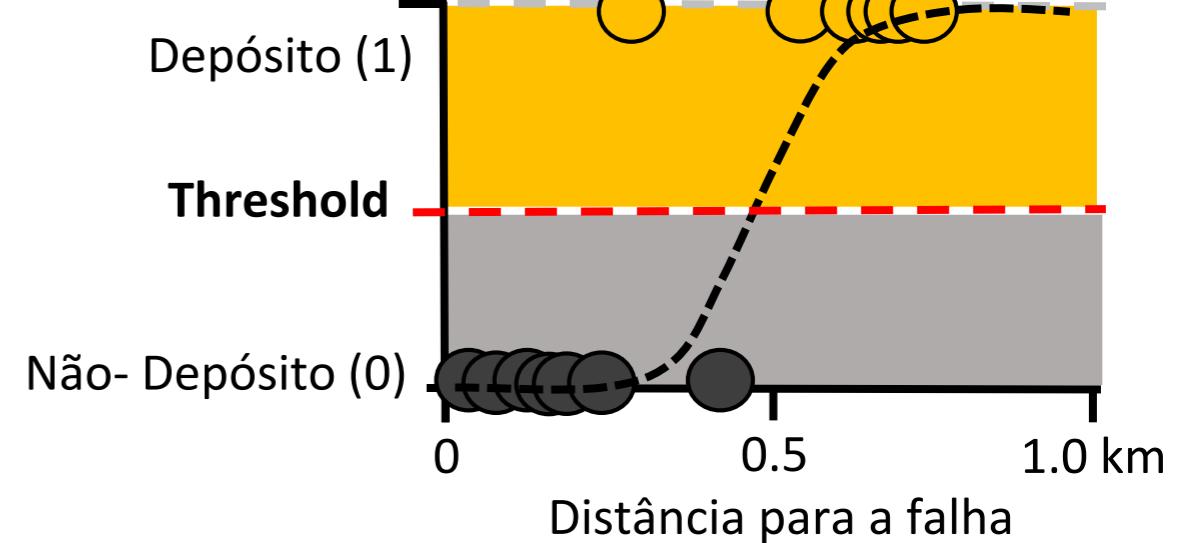
# Regressão Logística (Logistic Regression)

## Exemplo

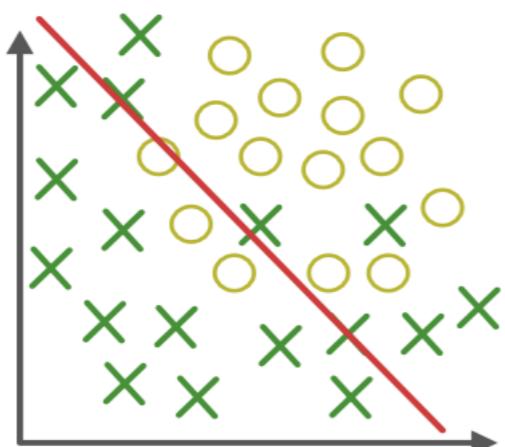


Depósito

Não – Depósito (estéril)

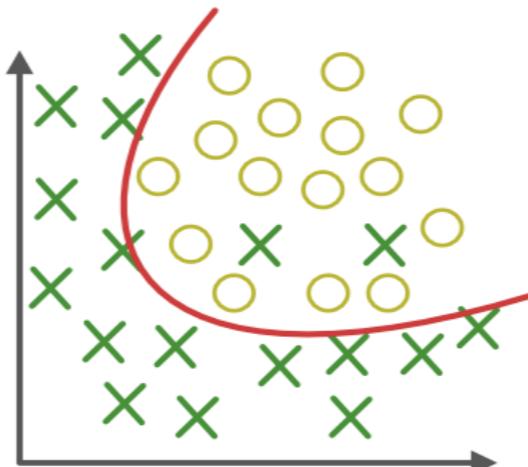


# Regularização



**Under-fitting**  
(too simple to  
explain the variance)

**High Bias!!**  
**Low Variance!!**



**Appropriate-fitting**

**Low Bias!!**  
**Low Variance!!**



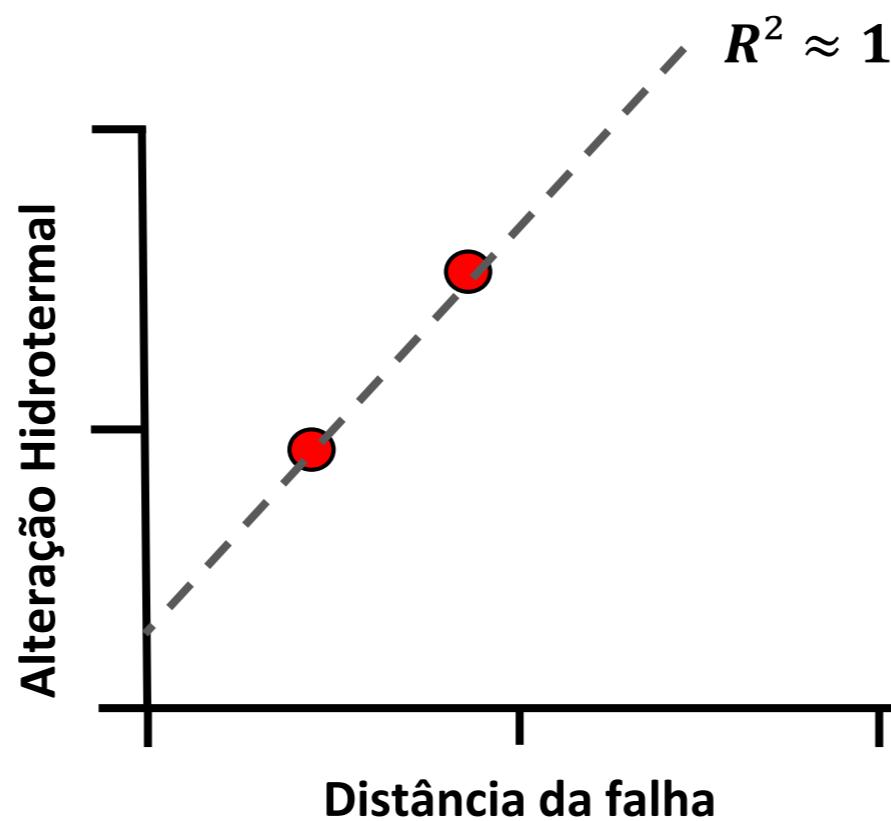
**Over-fitting**  
(forcefitting--too  
good to be true)

**Low Bias!!**  
**High Variance!!**

**Lasso Regression (L1)**

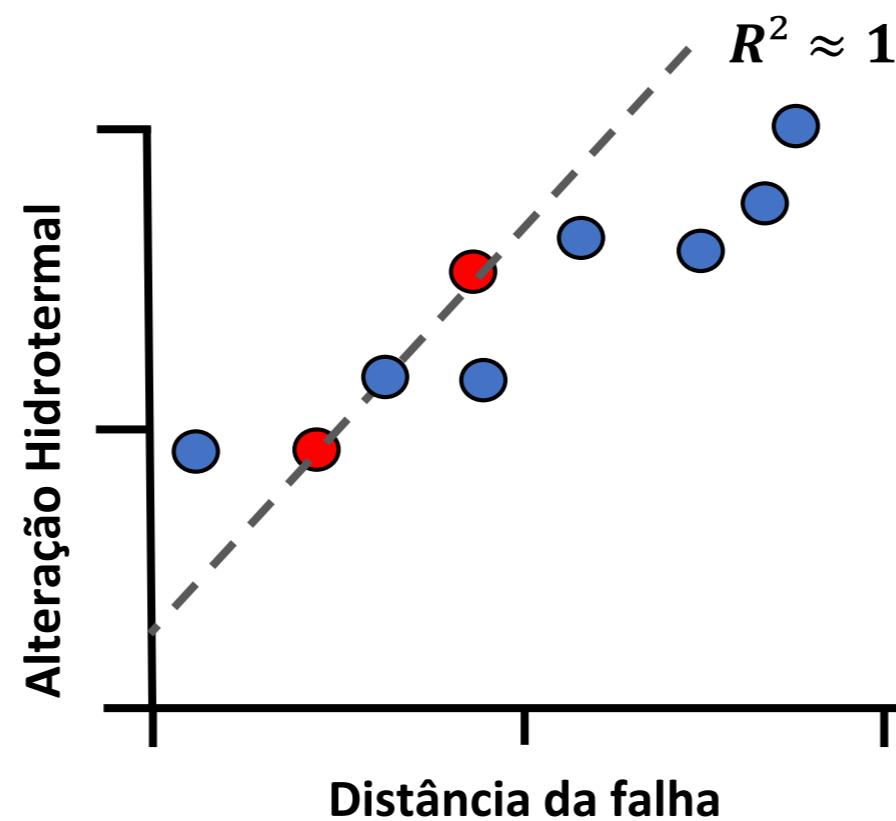
**Ridge Regression (L2)**

# Regularização



 Training Data

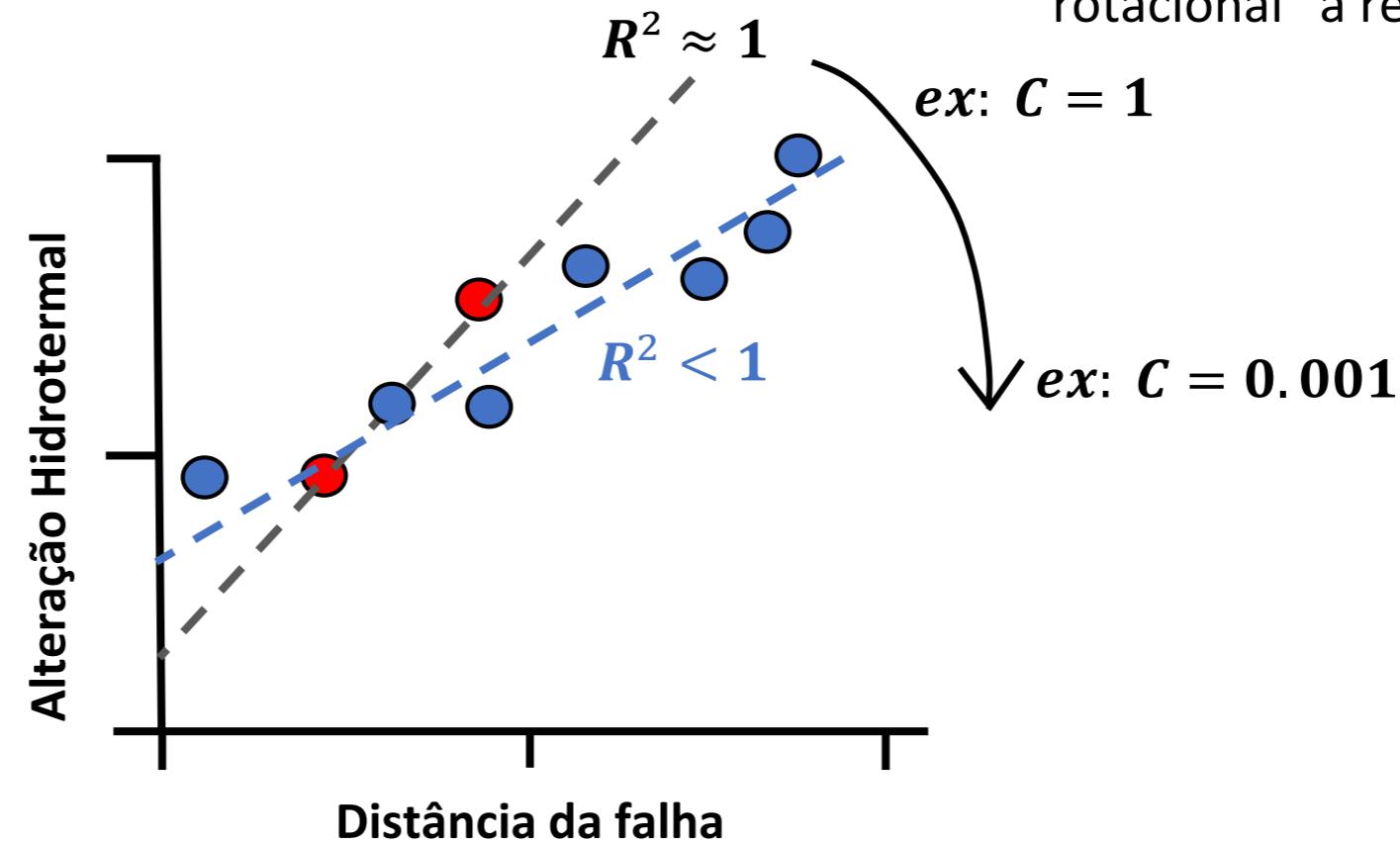
# Regularização



- Training Data
- Test Data

# Regularização

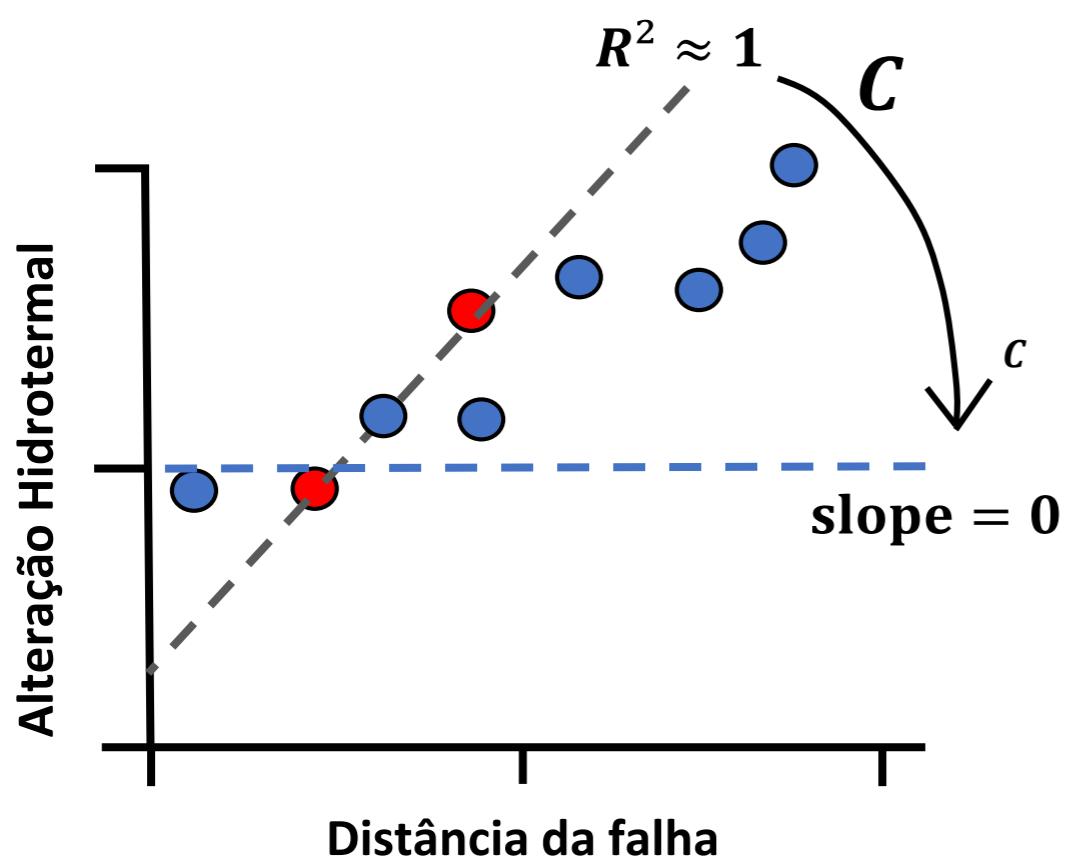
Para prevenir *overfitting*, nós inserimos um fator  $C$  que “rotacional” a reta ajustada.



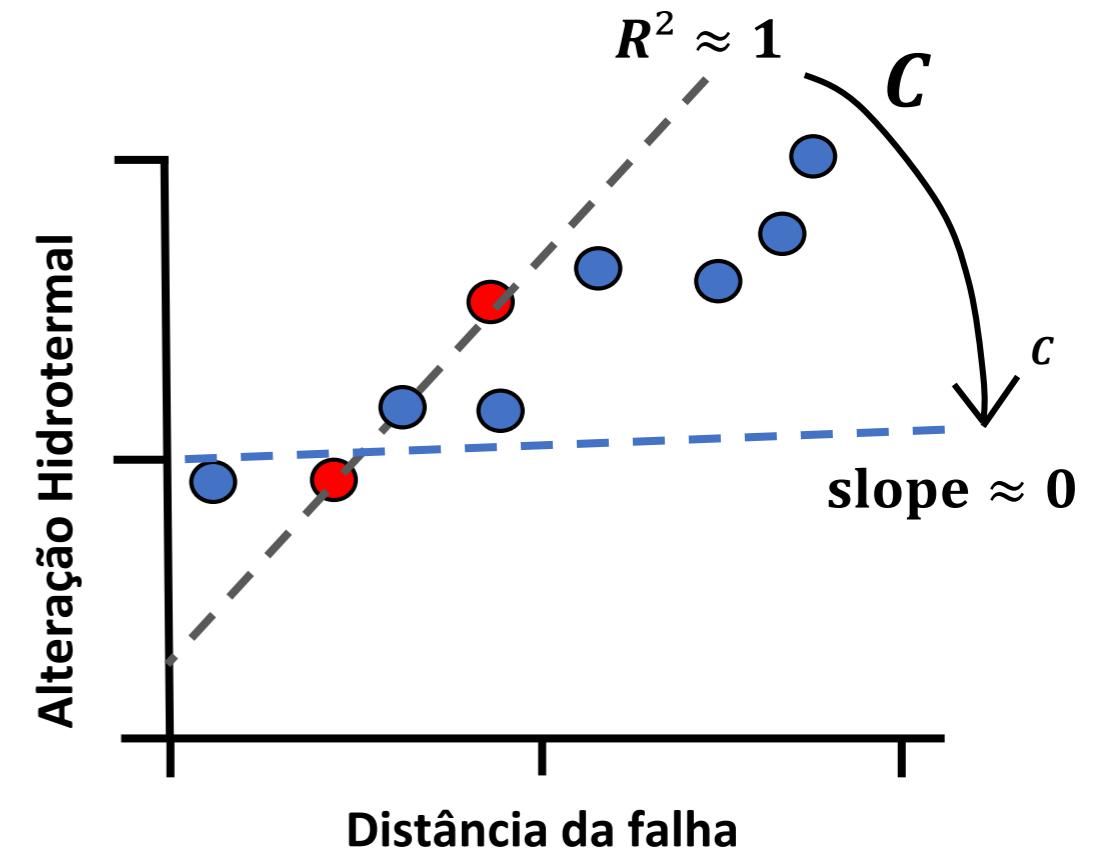
- Training Data
- Test Data

**Nota:**  $C = \frac{1}{\lambda}$

# Lasso Regression (L1)   vs   Ridge Regression (L2)

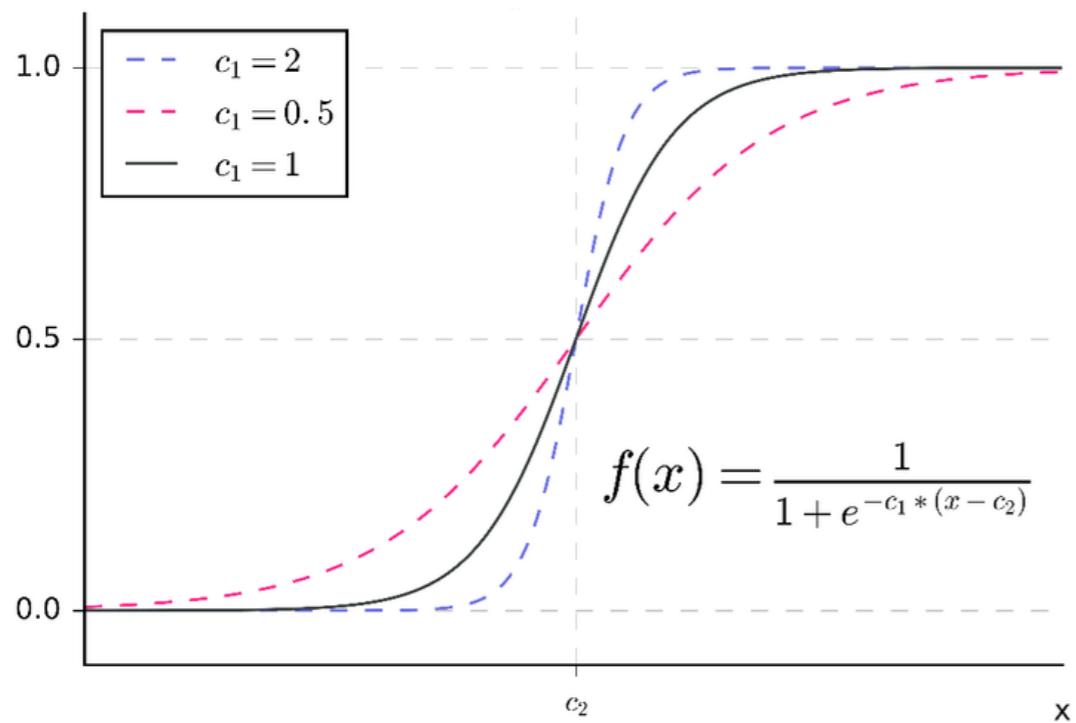


Lasso Regression (L1)

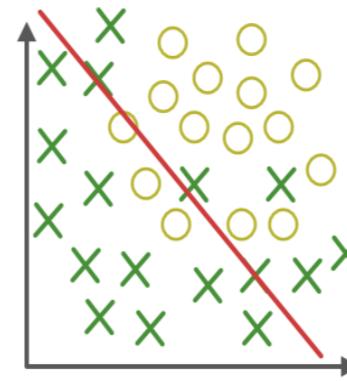


Ridge Regression (L2)

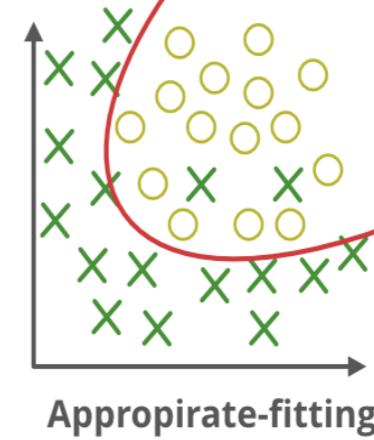
# Regularização



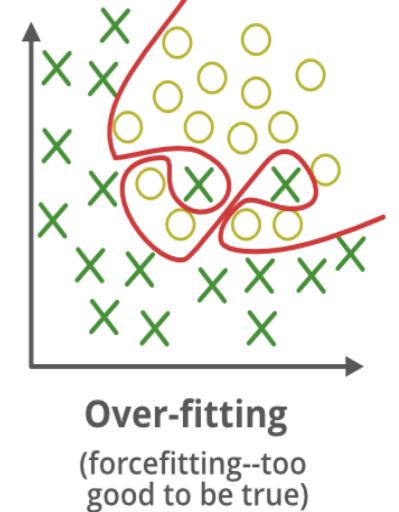
**$C$  pequeno**  
(alta regularização)



**$C$  ideal**  
(ótima regularização)



**$C$  grande**  
(baixa regularização)



# Regressão Logística na Geociências

**Reddy, R. K. T., & Bonham-Carter, G. F. (1991).** A decision-tree approach to mineral potential mapping in Snow Lake Area, Manitoba. *Canadian Journal of Remote Sensing*, 17, 191–200.

**Harris, D., & Pan, G. (1999).** Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis. *Natural Resources Research*, 8, 93– 109.

**Carranza, E.J.M., Hale, M. (2001).** Logistic regression for geologically-constrained mapping of gold mineralization potential, Baguio district, Philippines. *Exploration and Mining Geology Journal* 10(3), 165-175

**Harris, J.R., Sanborn-Barrie, M., Panagapko, D.A., Skulski, T., Parker, J.R. (2006).** Gold prospectivity maps of the Red Lake greenstone belt: application of GIS technology. *Can. J. Earth Sci.* 43, 865–893.

**Chen, C., Dai, H., Liu, Y., He, B. (2011).** Mineral Prospectivity Mapping Integrating Multi- source Geology Spatial Data Sets and Logistic Regression Modelling. pp. 214–217.

**Harris, D.P., Zurcher, L., Stanley, M., Marlow, J., Pan, G., (2003).** A comparative analysis of favourability mappings by weights of evidence, probabilistic neural networks, dis- criminant analysis, and logistic regression. *Nat. Resour. Res.* 12, 241–255.

# Regressão Logística na Geociências

Reddy, R. K. T., & Bonham-Carter, G. F. (1991). A decision-tree approach to mineral potential mapping in Snow Lake Area, Manitoba. *Canadian Journal of Remote Sensing*, 17, 191–200.

Harris, D., & Pan, G. (1999). Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and decision trees. *Geological Society of America Special Paper*, 340, 1–22.

Harris, D.P., Zurcher, E.J., and Harris, J.R. (2000). Mineral favorability mappings by weight-of-evidence logistic regression. *Nat. Resources Res.* 9, 111–125.

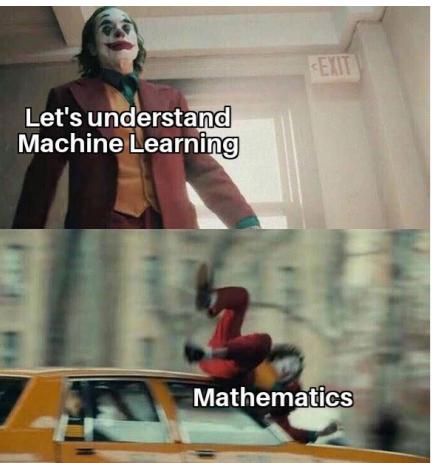
Harris, J.R., Sanborn, M., and Zurcher, E.J. (2002). Mineral favorability maps of the Red Lake greenstone belt, Ontario, Canada. *Geological Society of America Special Paper*, 364, 1–22.

**As aplicações de Regressão logística na Geociências, em sua maioria, negligenciam a importância da regularização!!**

Chen, C., Dai, H., Liu, Y., He, B. (2011). Mineral Prospectivity Mapping Integrating Multi- source Geology Spatial Data Sets and Logistic Regression Modelling. pp. 214–217.

Carranza, E.J.M., Hale, M. (2001). Logistic regression for geologically-constrained mapping of gold mineralization potential, Baguio district, Philippines. *Exploration and Mining Geology Journal* 10, 171–175.

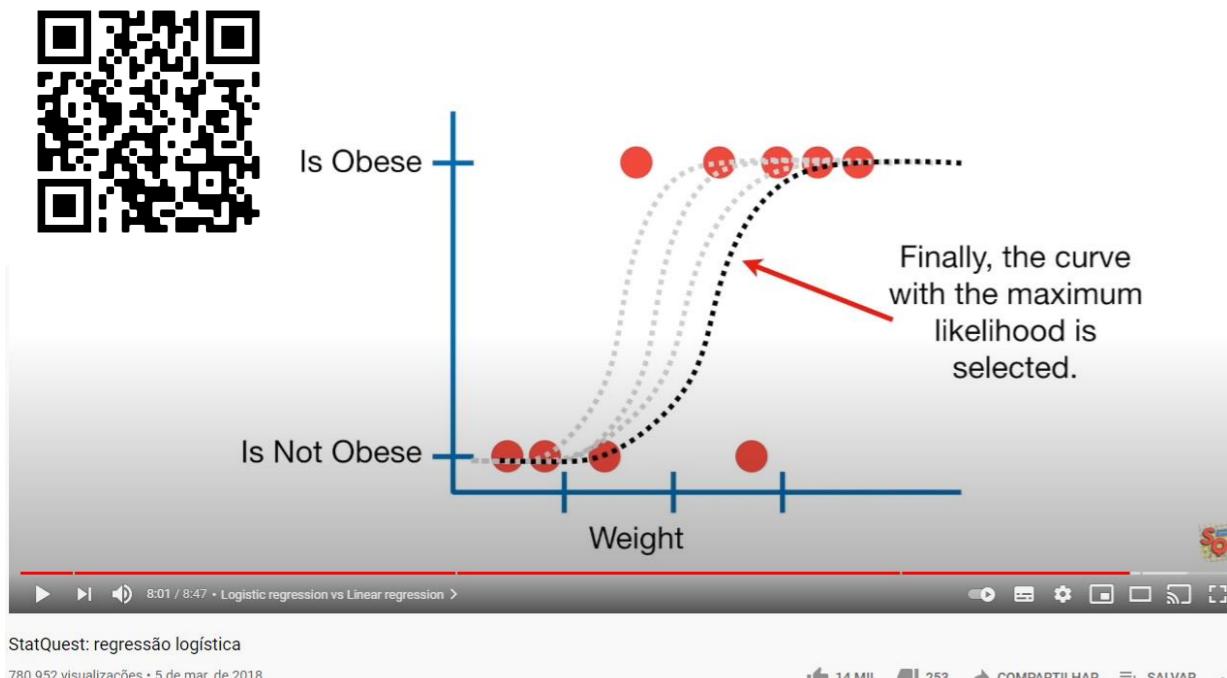




# Regressão Logística (Logistic Regression)

*Want to know more??*

**Easy mode**



**StatQuest with Josh Starmer**

<https://www.youtube.com/watch?v=yIYKR4sgzl8>

**Math and statistics required!!**

Logistic Regression

- Predicts the probability of poor care
  - Denote dependent variable "PoorCare" by  $y$
  - $P(y = 1)$
- Then  $P(y = 0) = 1 - P(y = 1)$
- Independent variables  $x_1, x_2, \dots, x_k$
- Uses the Logistic Response Function

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}$$

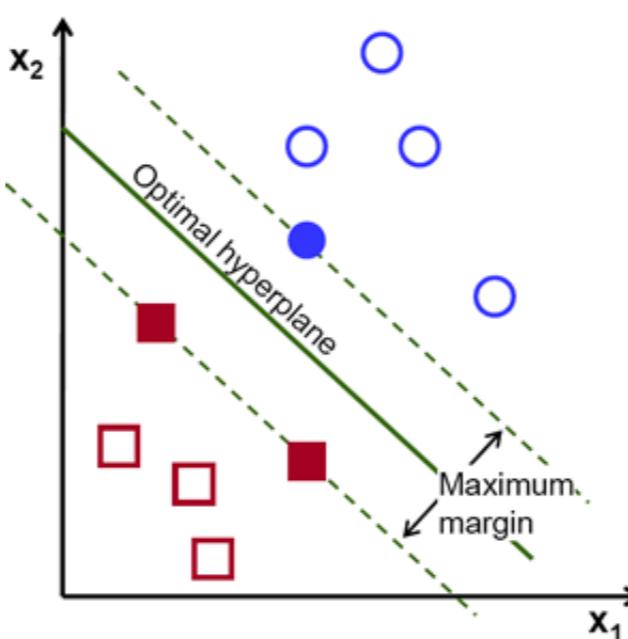
Nonlinear transformation of linear regression equation to produce number between 0 and 1

15.071x - Modeling the Expert: An Introduction to Logistic Regression  
3.2.4 Introduction to Logistical Regression - Video 3: Logistic Regression  
2.908 visualizações • 13 de dez. de 2018  
14 1 COMPARLHAR SALVAR ...

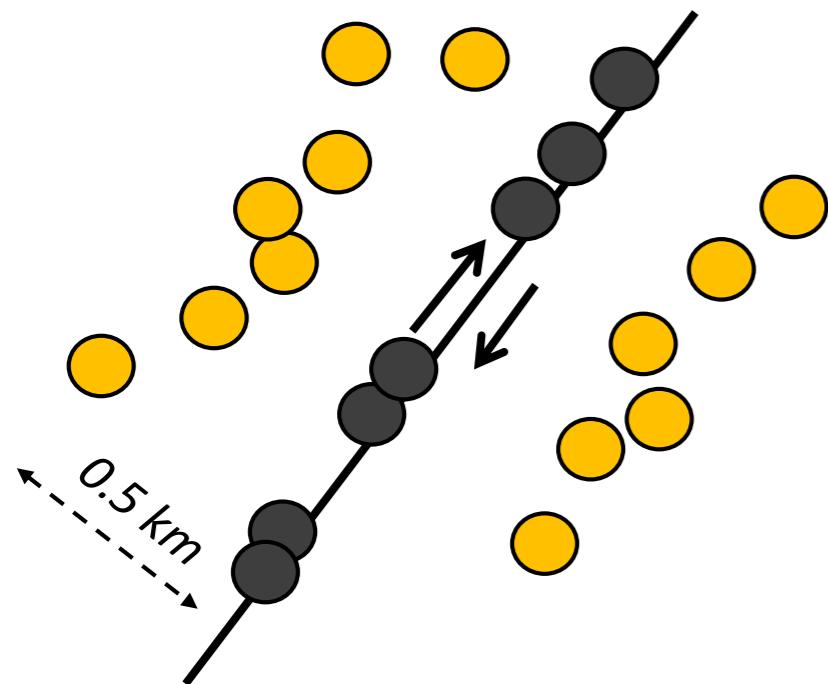
**MIT OpenCourseWare**

<https://www.youtube.com/watch?v=U57wvHVpe-8>

# Support Vector Machines (SVM)



# SUPPORT VECTOR MACHINES (SVM)



Depósito

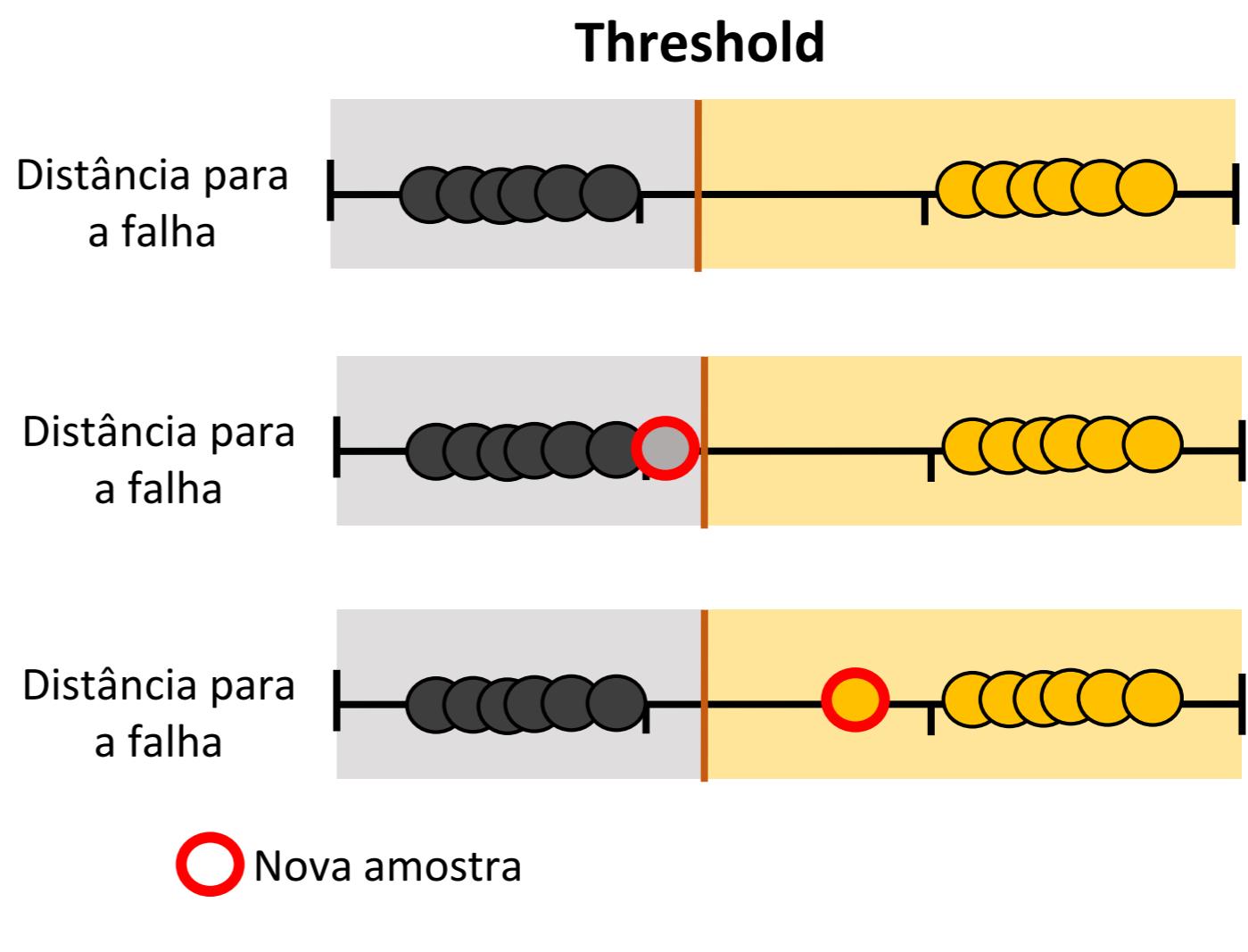
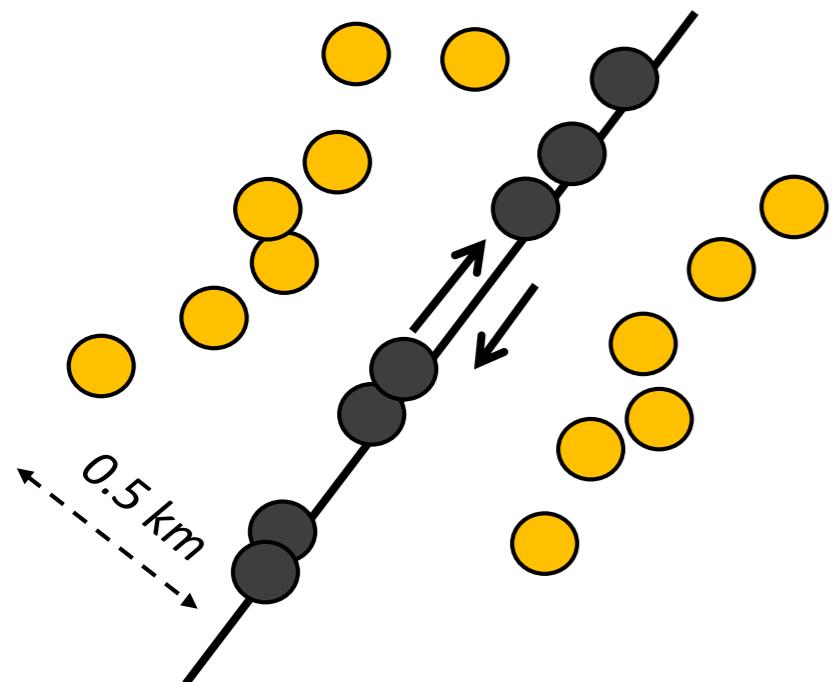
Não – Depósito (estéril)

Distância para  
a falha



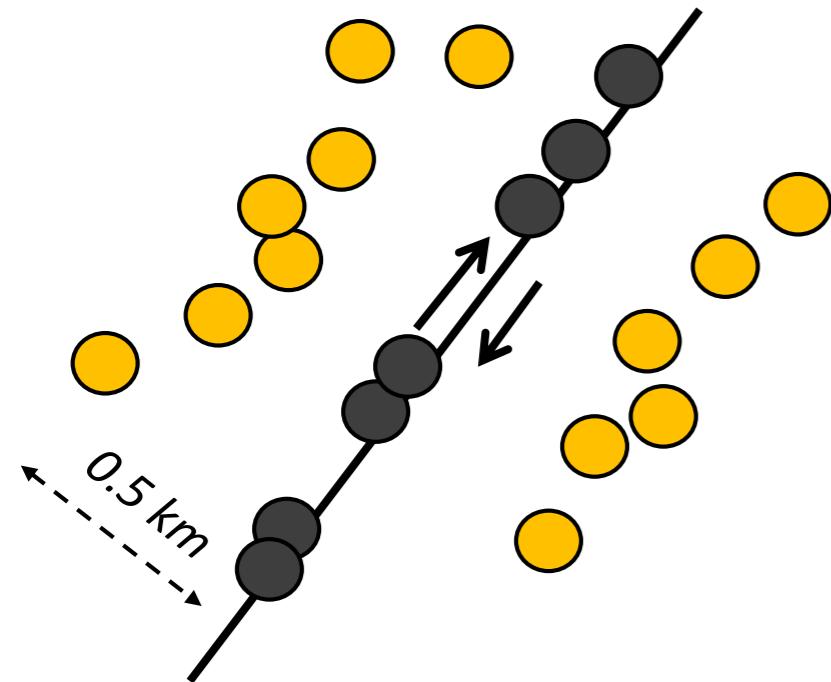
# SUPPORT VECTOR MACHINES (SVM)

## Exemplo

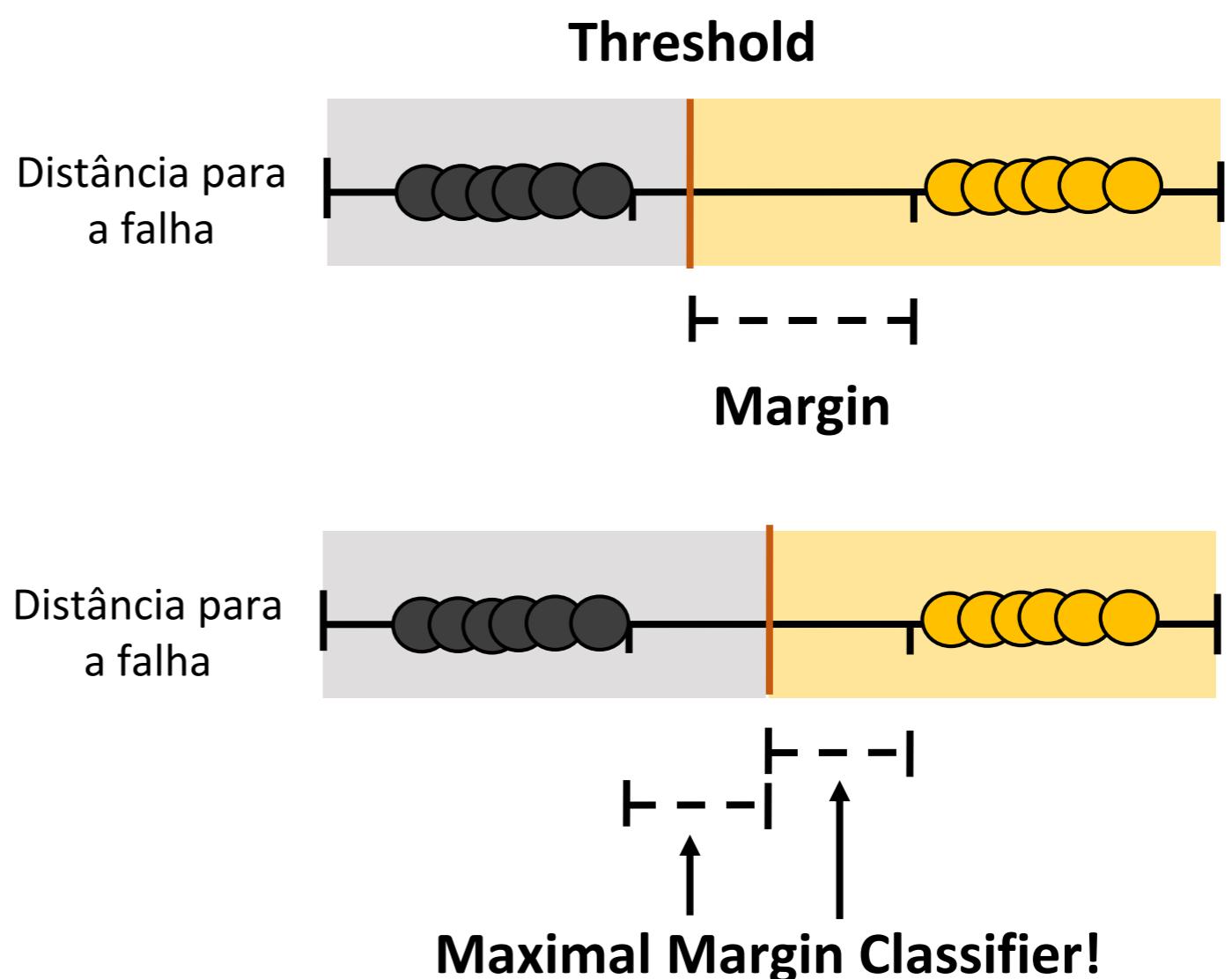


# SUPPORT VECTOR MACHINES (SVM)

## Exemplo

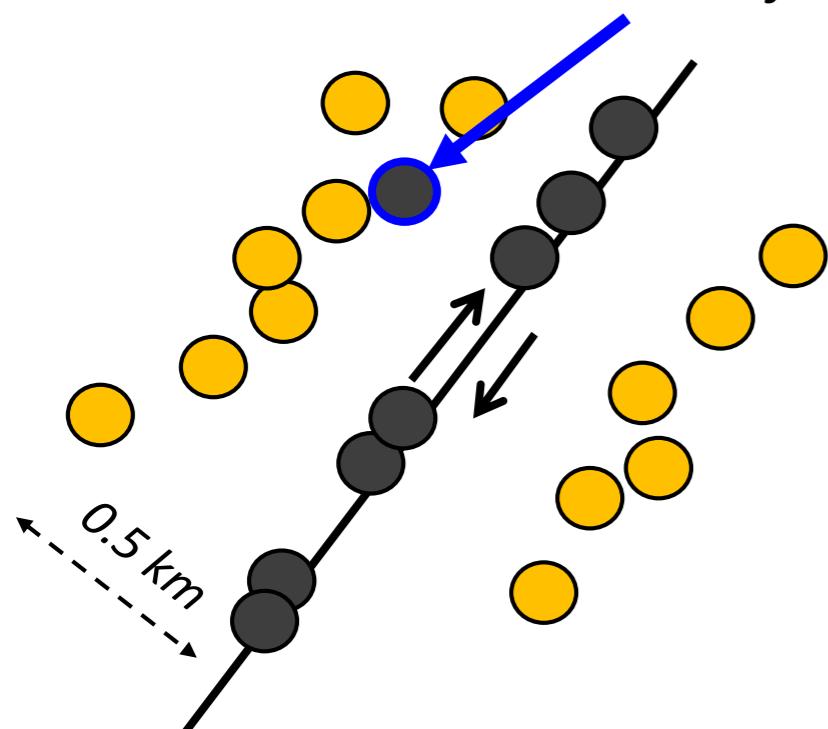


- Depósito
- Não – Depósito (estéril)



# SUPPORT VECTOR MACHINES (SVM)

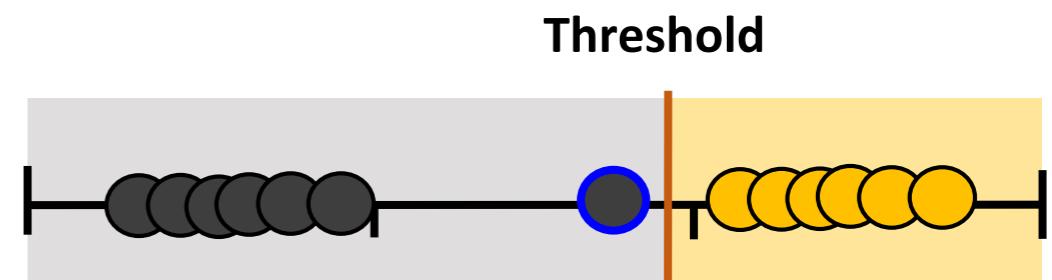
## *Outlier or missclassification*



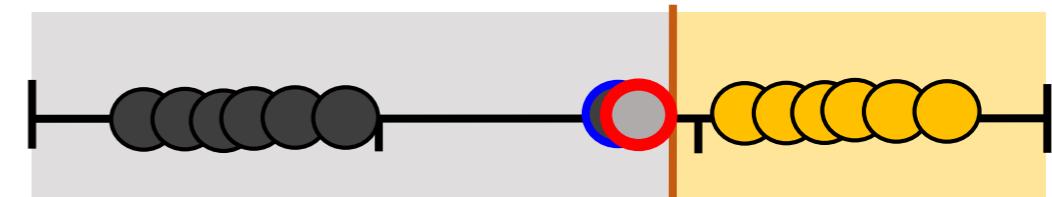
- Depósito
- Não – Depósito (estéril)

Distância para  
a falha

Distância para  
a falha



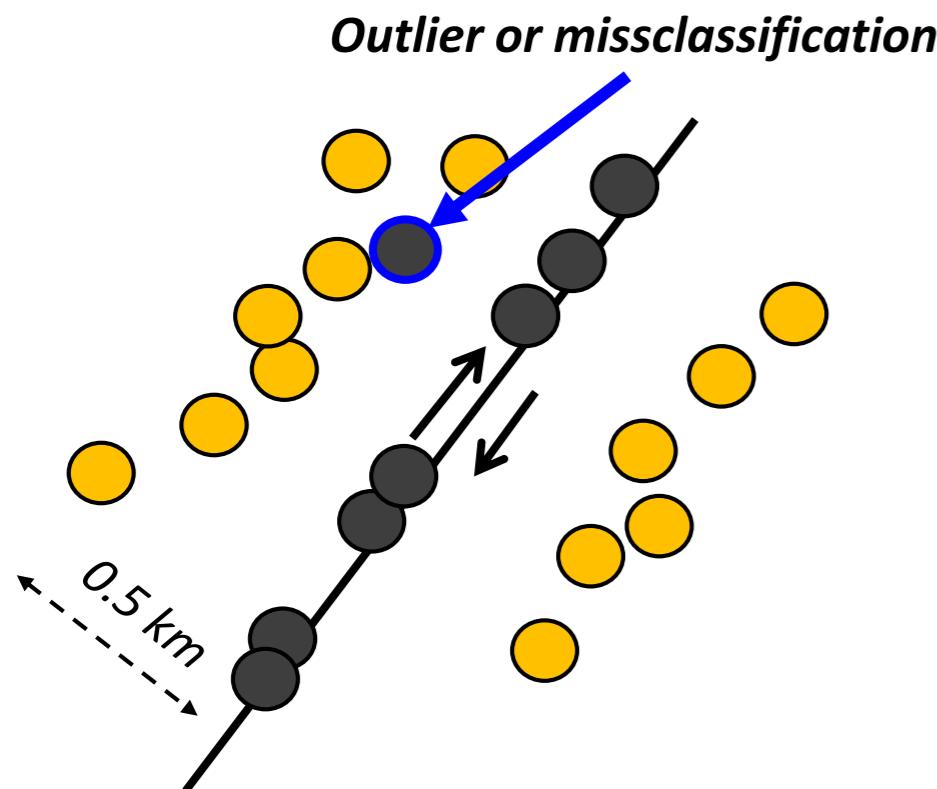
Nova amostra **Bad Threshold!**



*Para que o threshold não seja suscetível a “outliers”,  
é necessário que o algoritmo permita  
“missclassifications”*

# SUPPORT VECTOR MACHINES (SVM)

## Exemplo

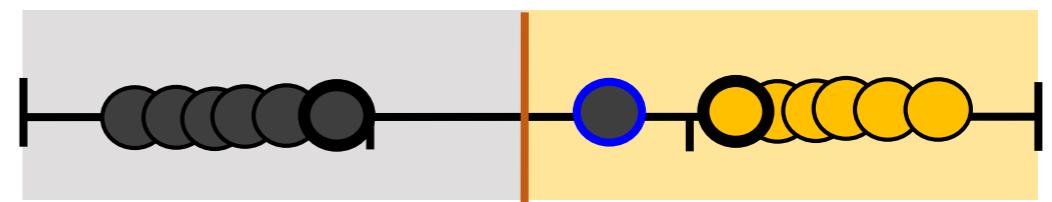


Depósito

Não – Depósito (estéril)

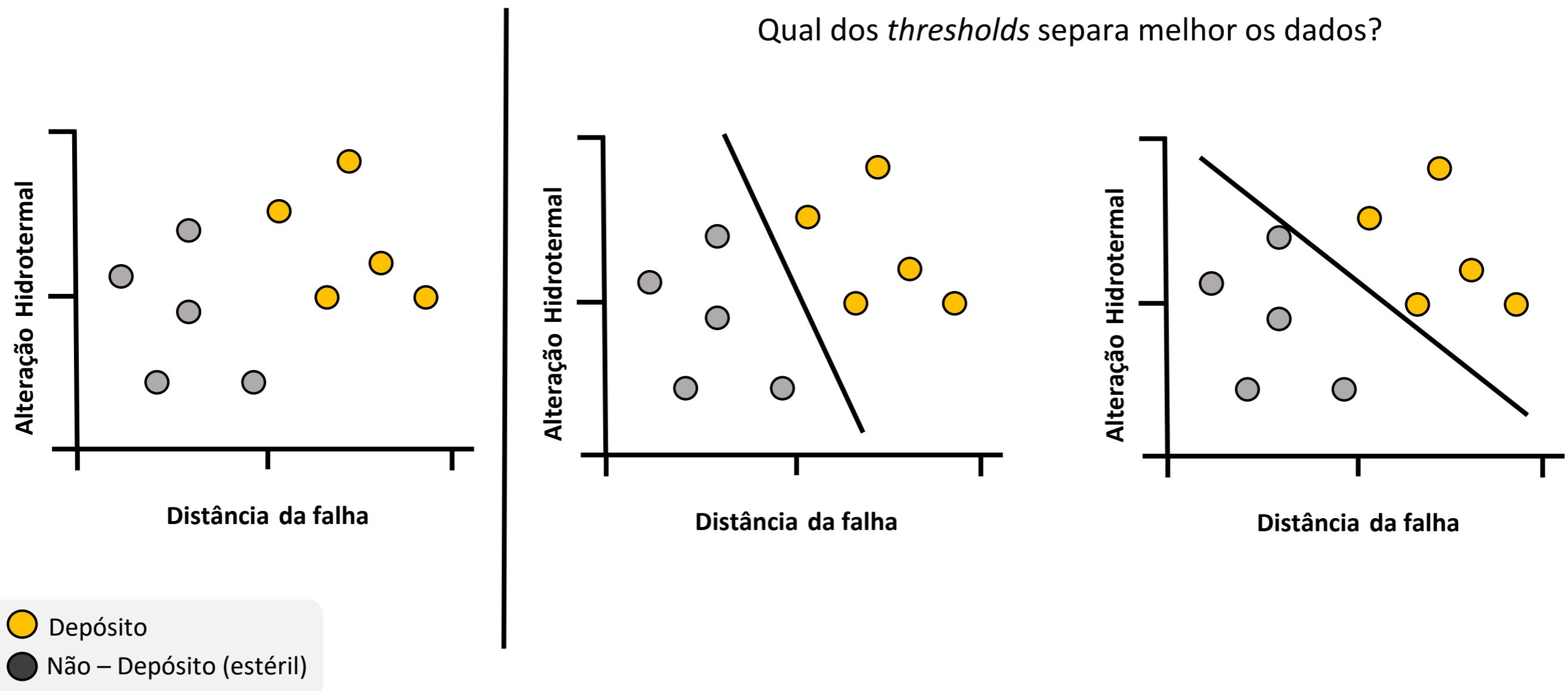
Distância para  
a falha

*Threshold que permite “missclassifications”*



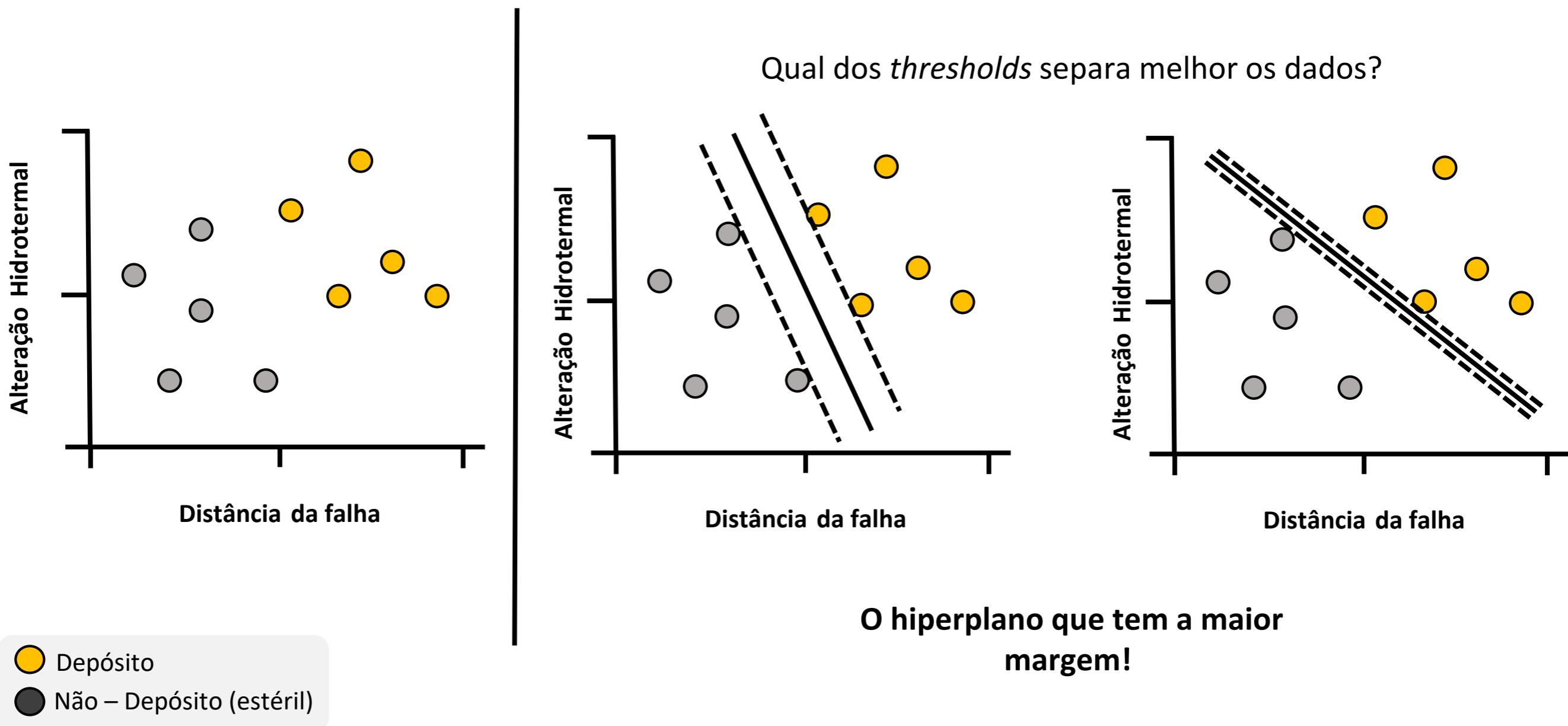
# SUPPORT VECTOR MACHINES (SVM)

## Exemplo 2D



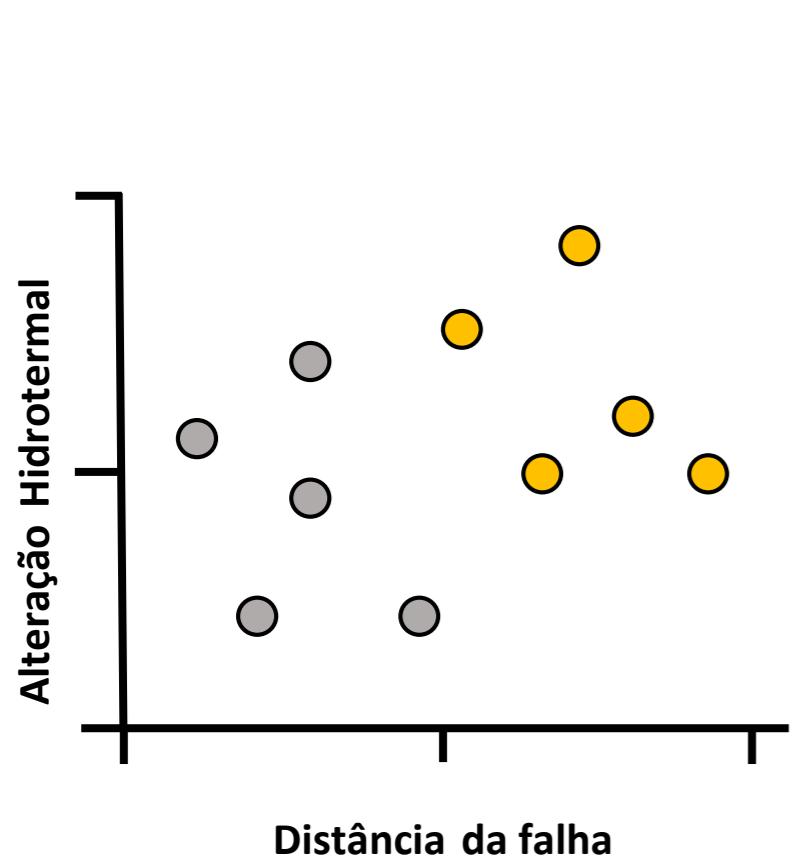
# SUPPORT VECTOR MACHINES (SVM)

## Exemplo 2D

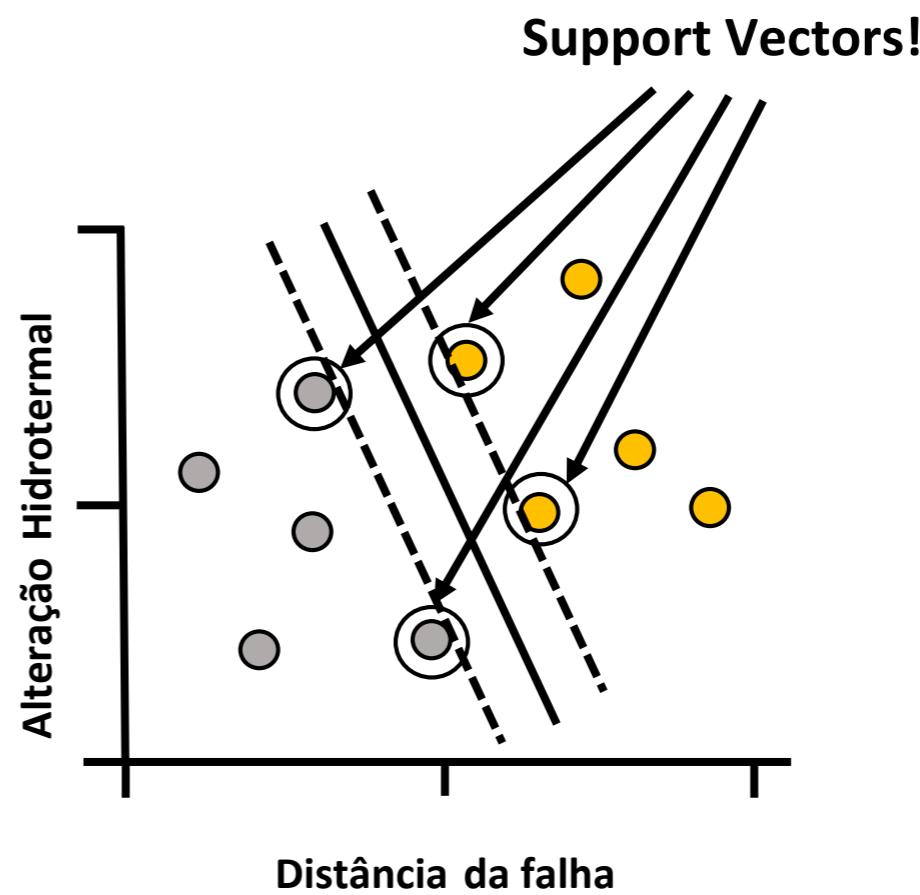


# SUPPORT VECTOR MACHINES (SVM)

## Exemplo 2D

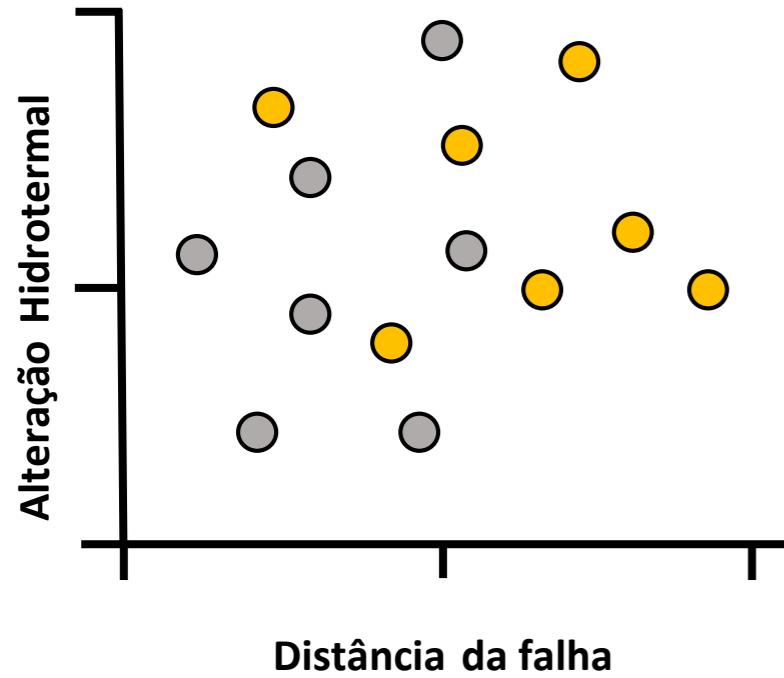


- Depósito (Yellow circle)
- Não – Depósito (estéril) (Grey circle)



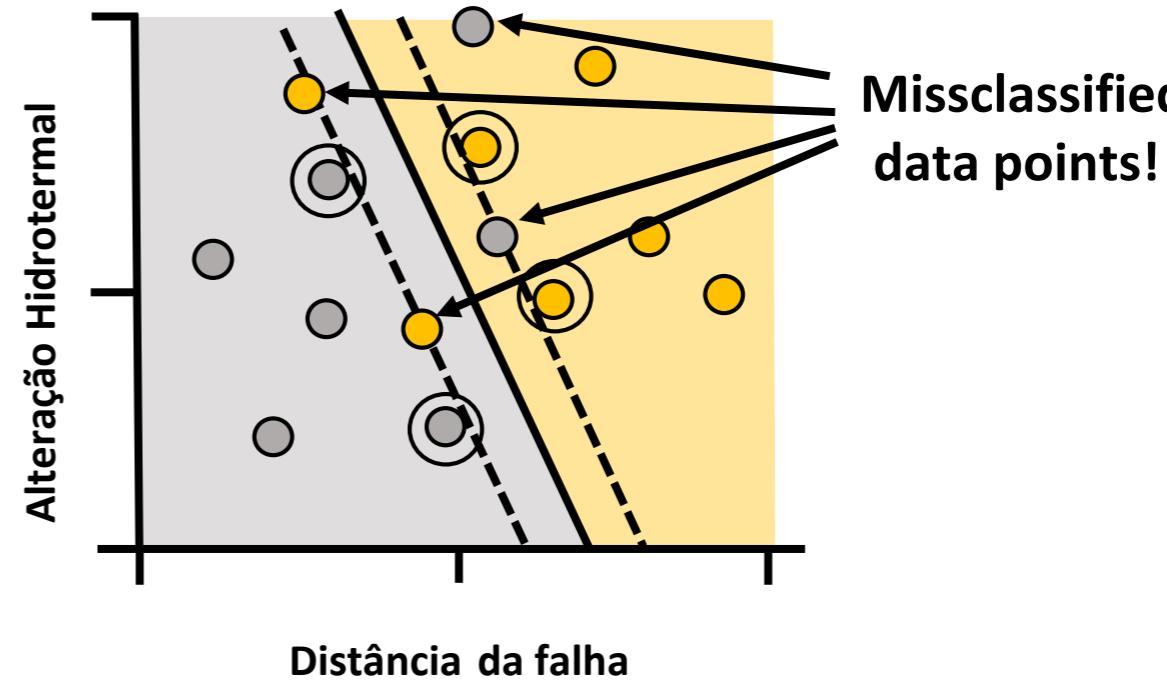
# SUPPORT VECTOR MACHINES (SVM)

## Exemplo 2D



- Depósito
- Não – Depósito (estéril)

Mas e se houver **missclassified data points**?

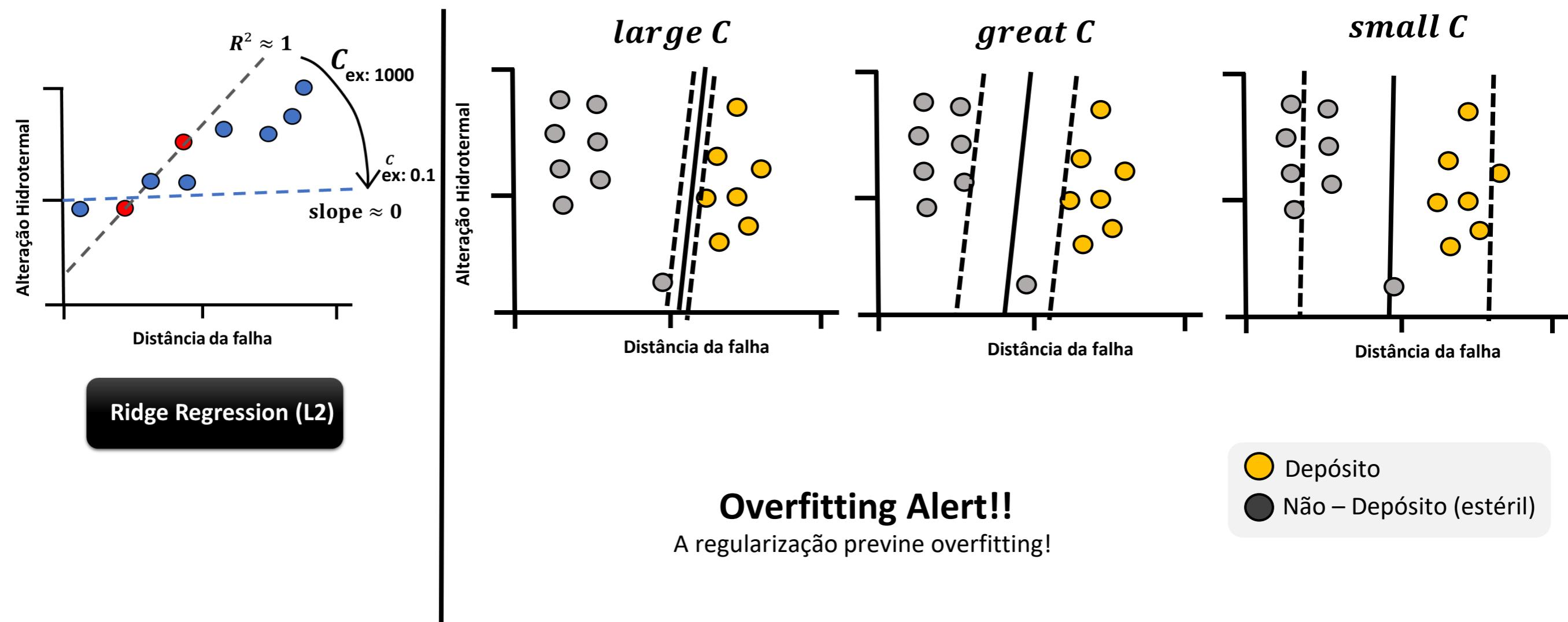


Missclassified data points!



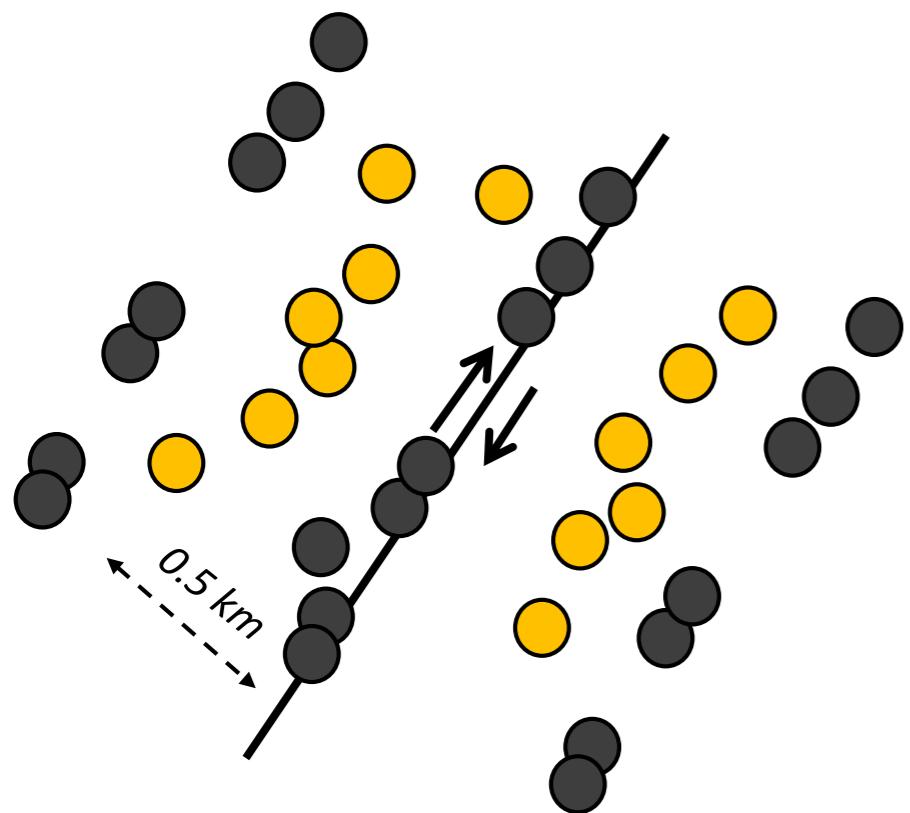
# Regularização

## *Ridge Regression (L2)*

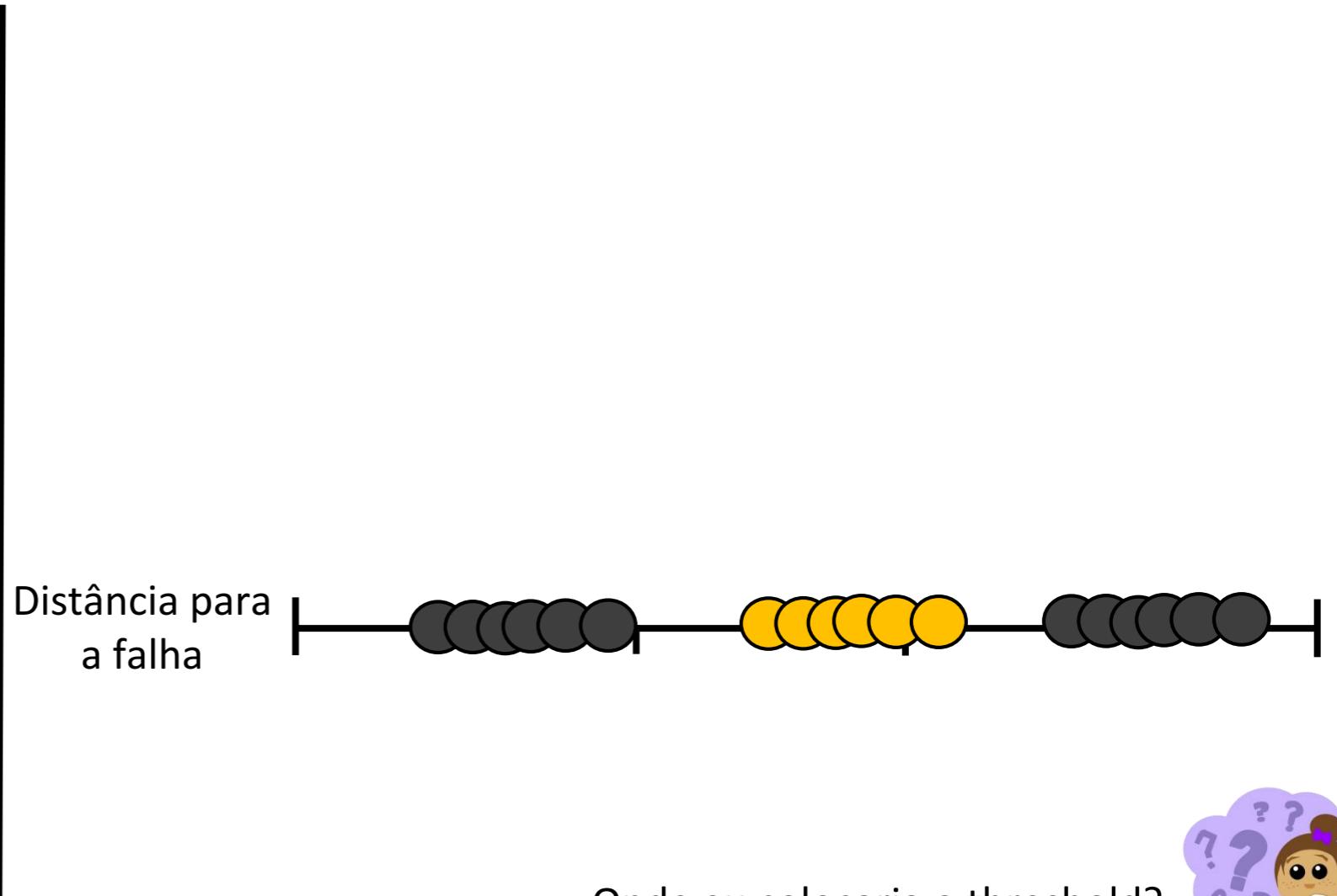


# NON-LINEAR SUPPORT VECTOR MACHINES (SVM)

## Exemplo



- Depósito
- Não – Depósito (estéril)

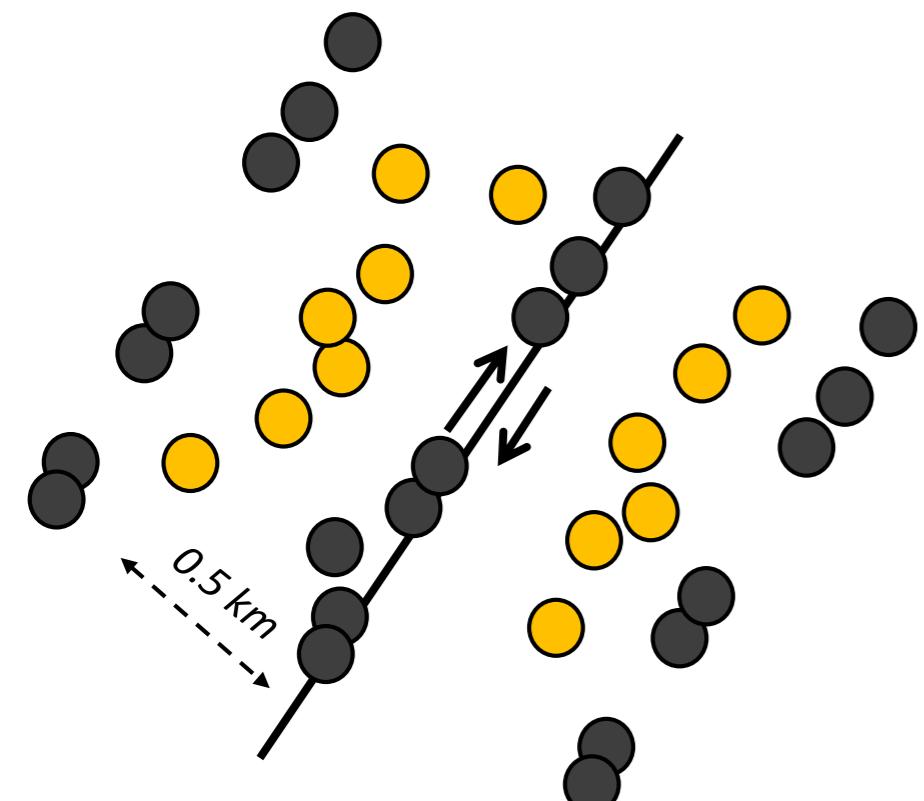


Onde eu colocaria o threshold?



# NON-LINEAR SUPPORT VECTOR MACHINES (SVM)

## Exemplo 1D

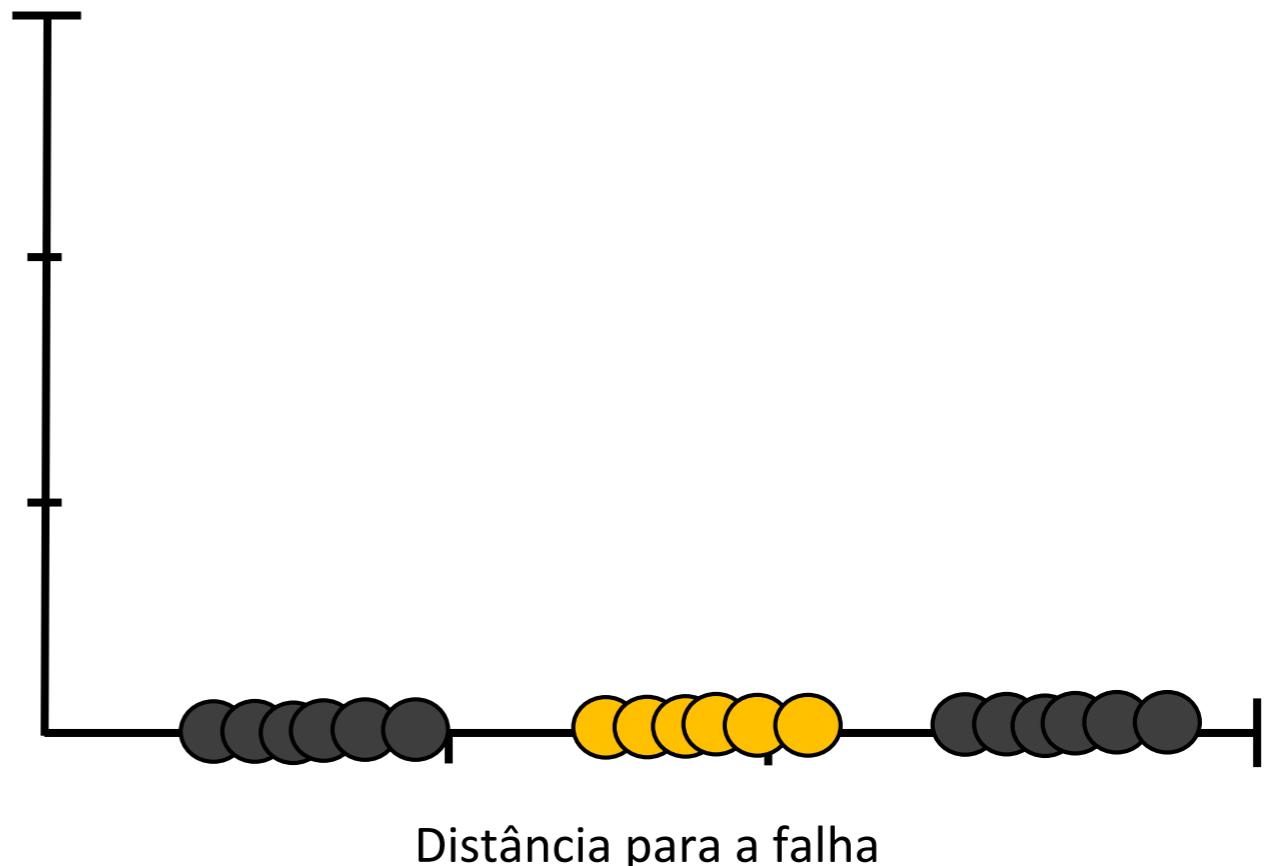


Depósito

Não – Depósito (estéril)

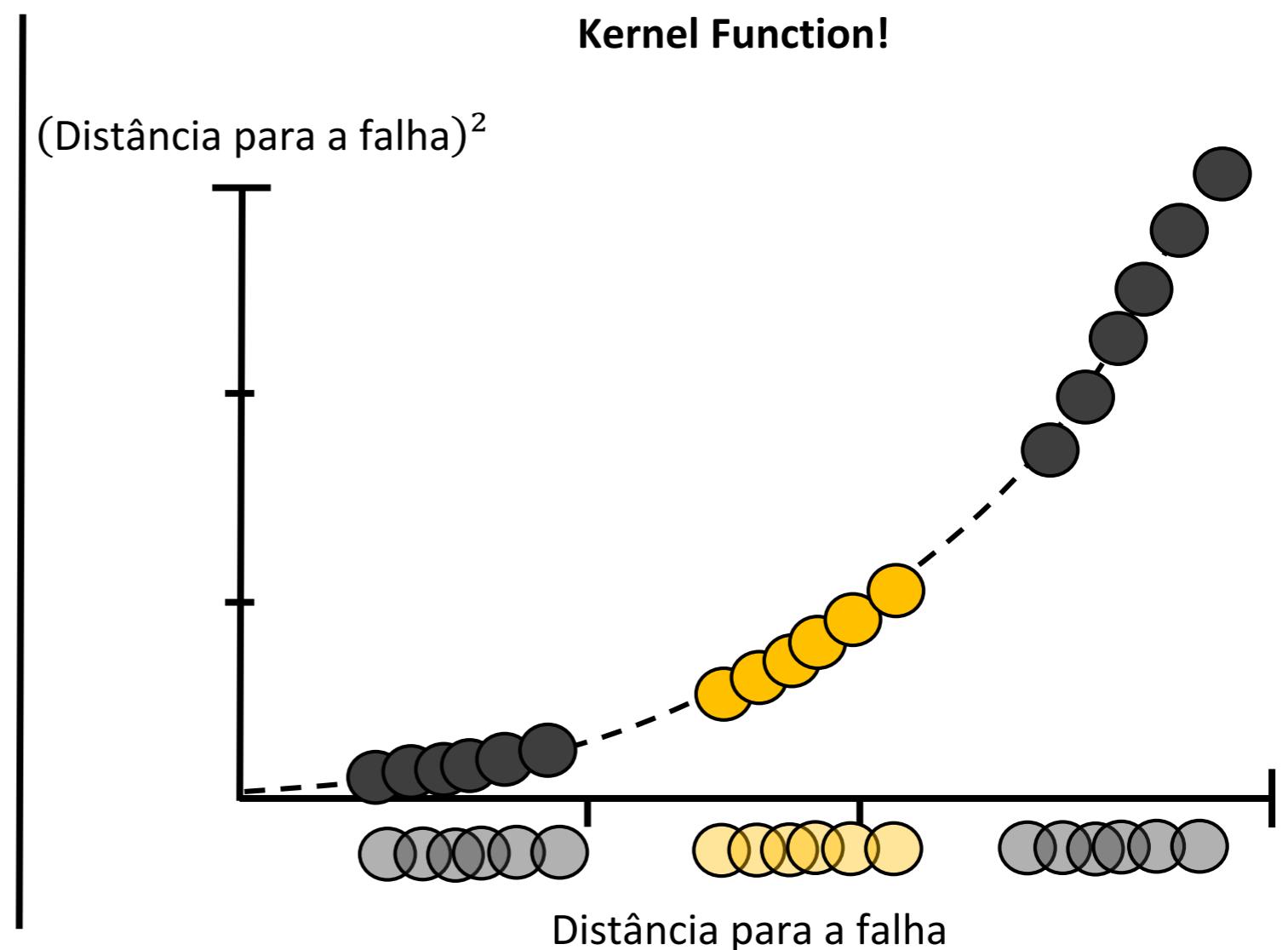
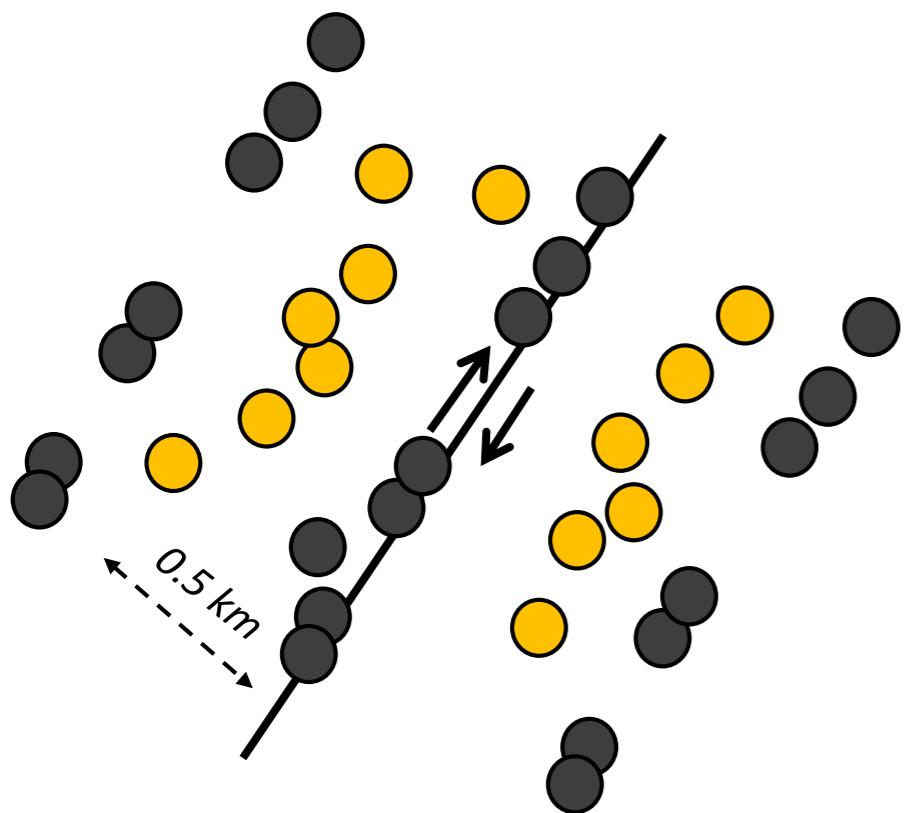
## Kernel Function!

$(\text{Distância para a falha})^2$



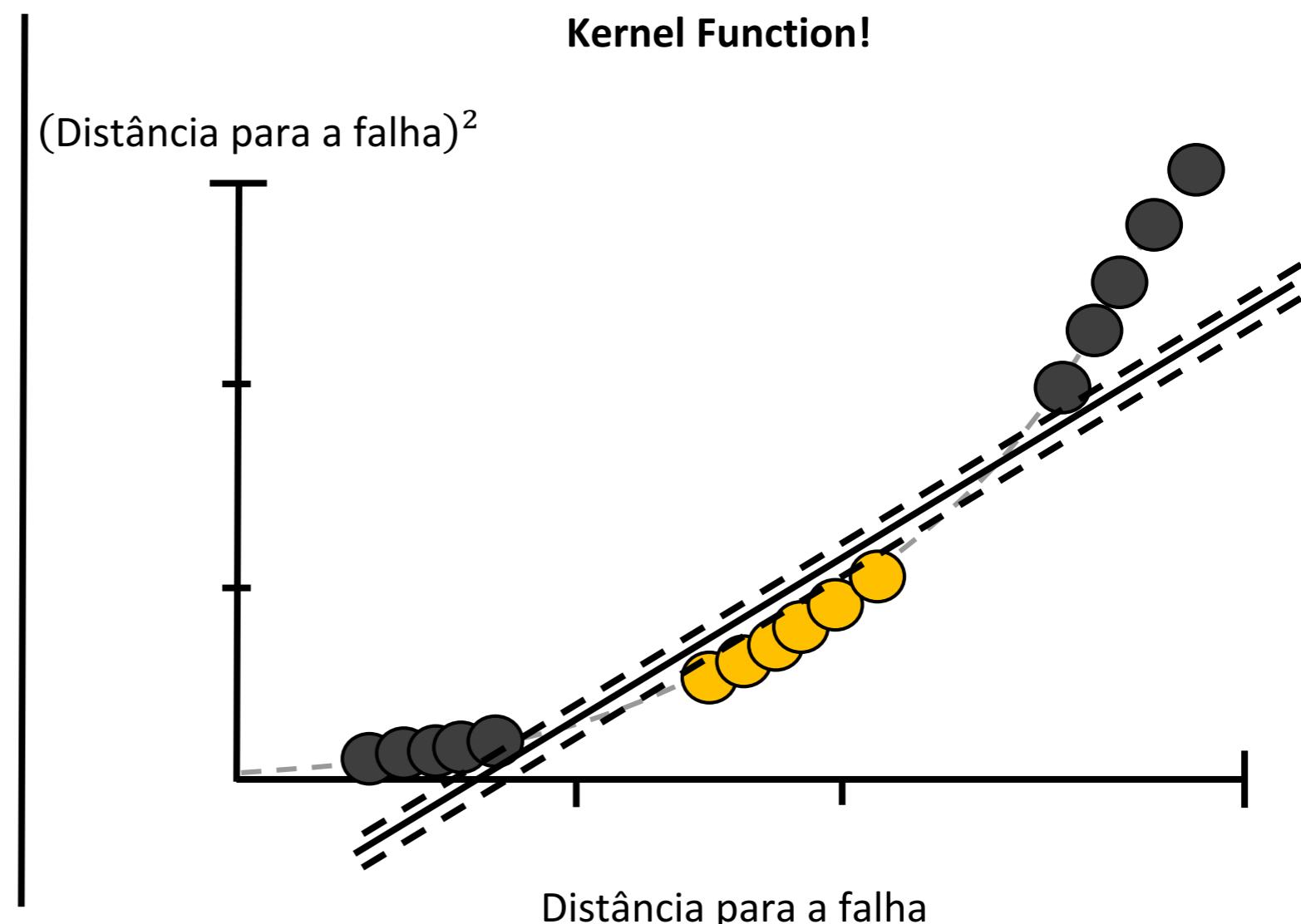
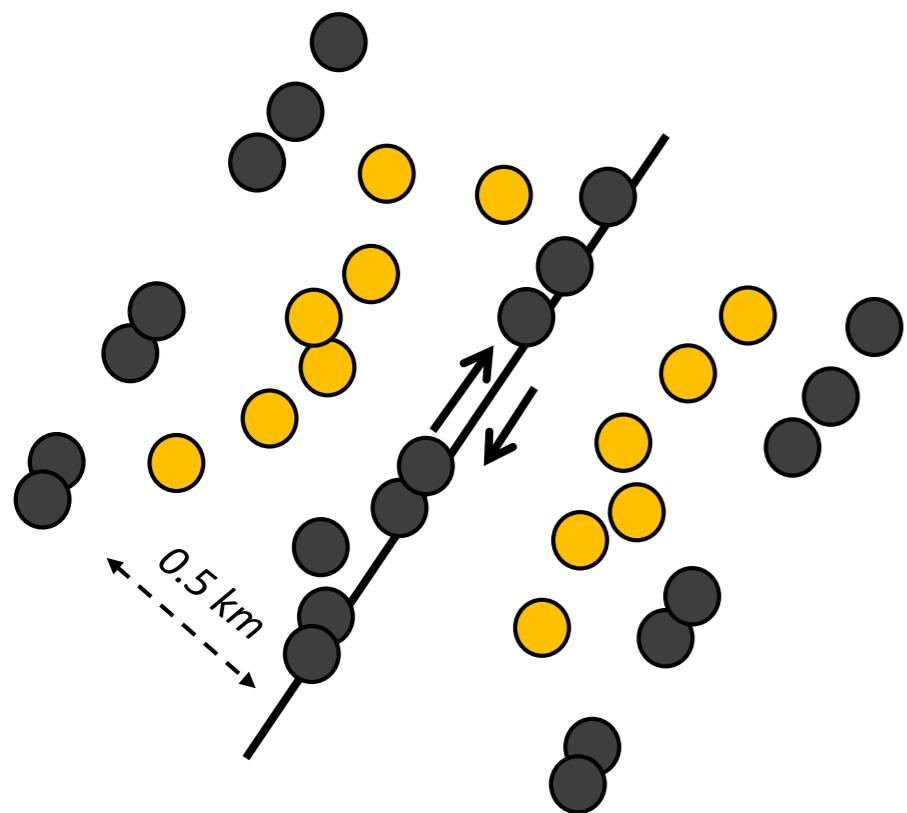
# NON-LINEAR SUPPORT VECTOR MACHINES (SVM)

## Exemplo 1D

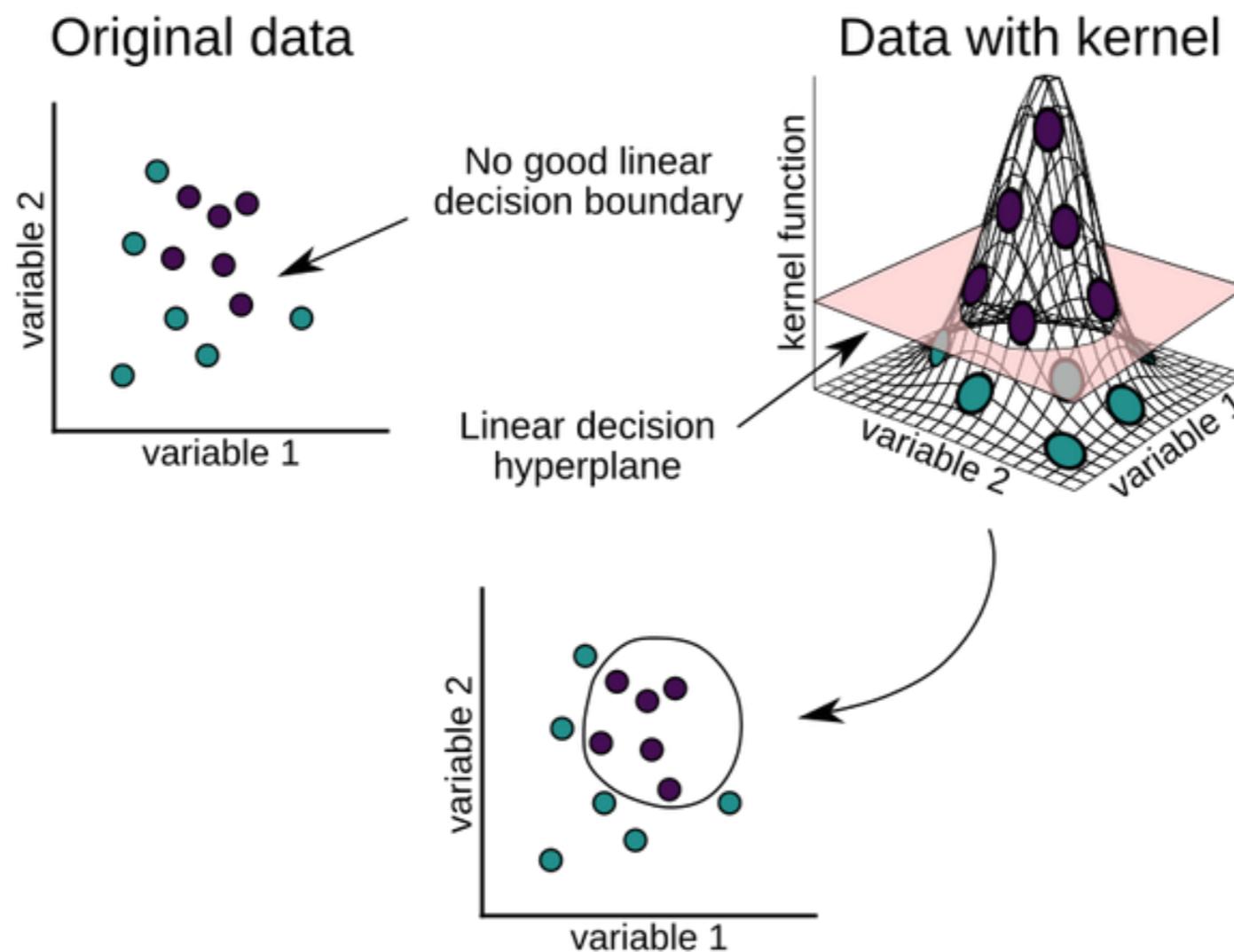


# NON-LINEAR SUPPORT VECTOR MACHINES (SVM)

## Exemplo 1D

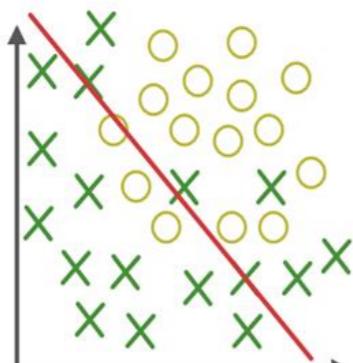


# NON-LINEAR SUPPORT VECTOR MACHINES (SVM)



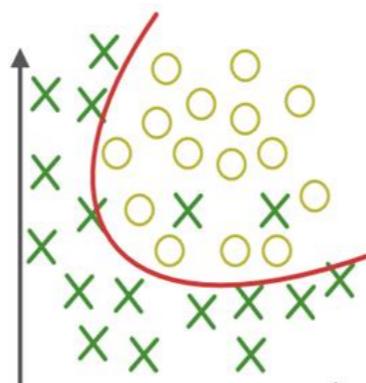
# SVM

**$C$  pequeno**  
(alta regularização)



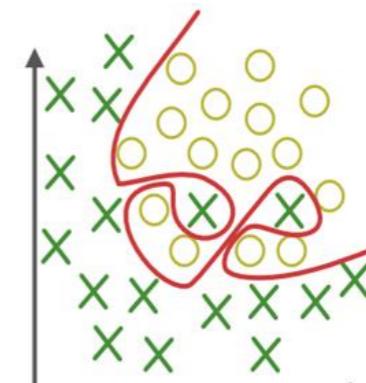
**Under-fitting**  
(too simple to  
explain the variance)

**$C$  ideal**  
(ótima regularização)

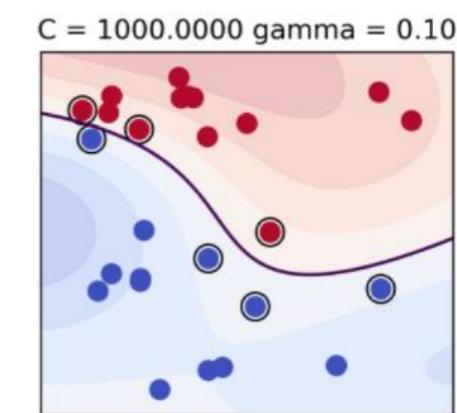
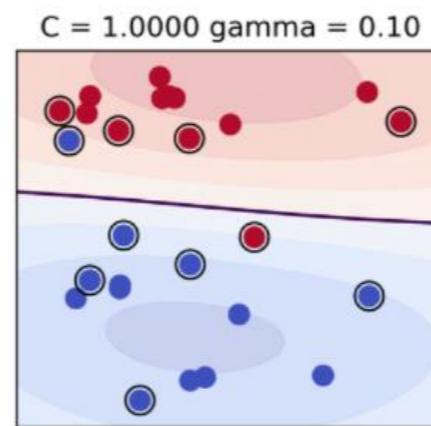
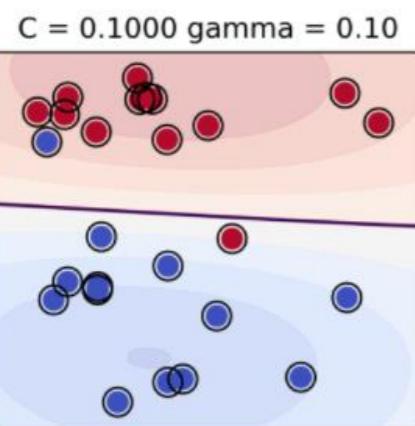


**Appropriate-fitting**

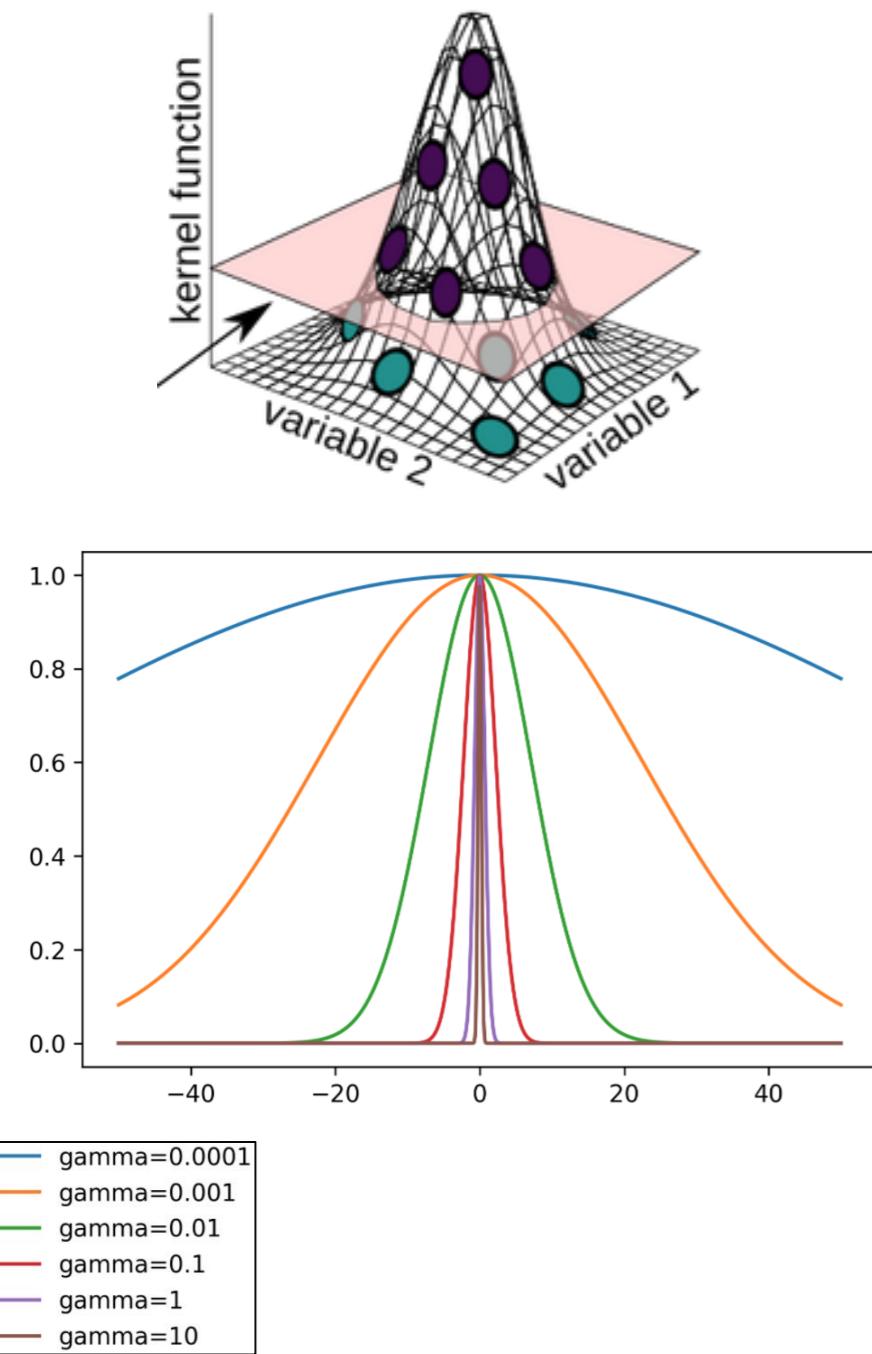
**$C$  grande**  
(baixa regularização)



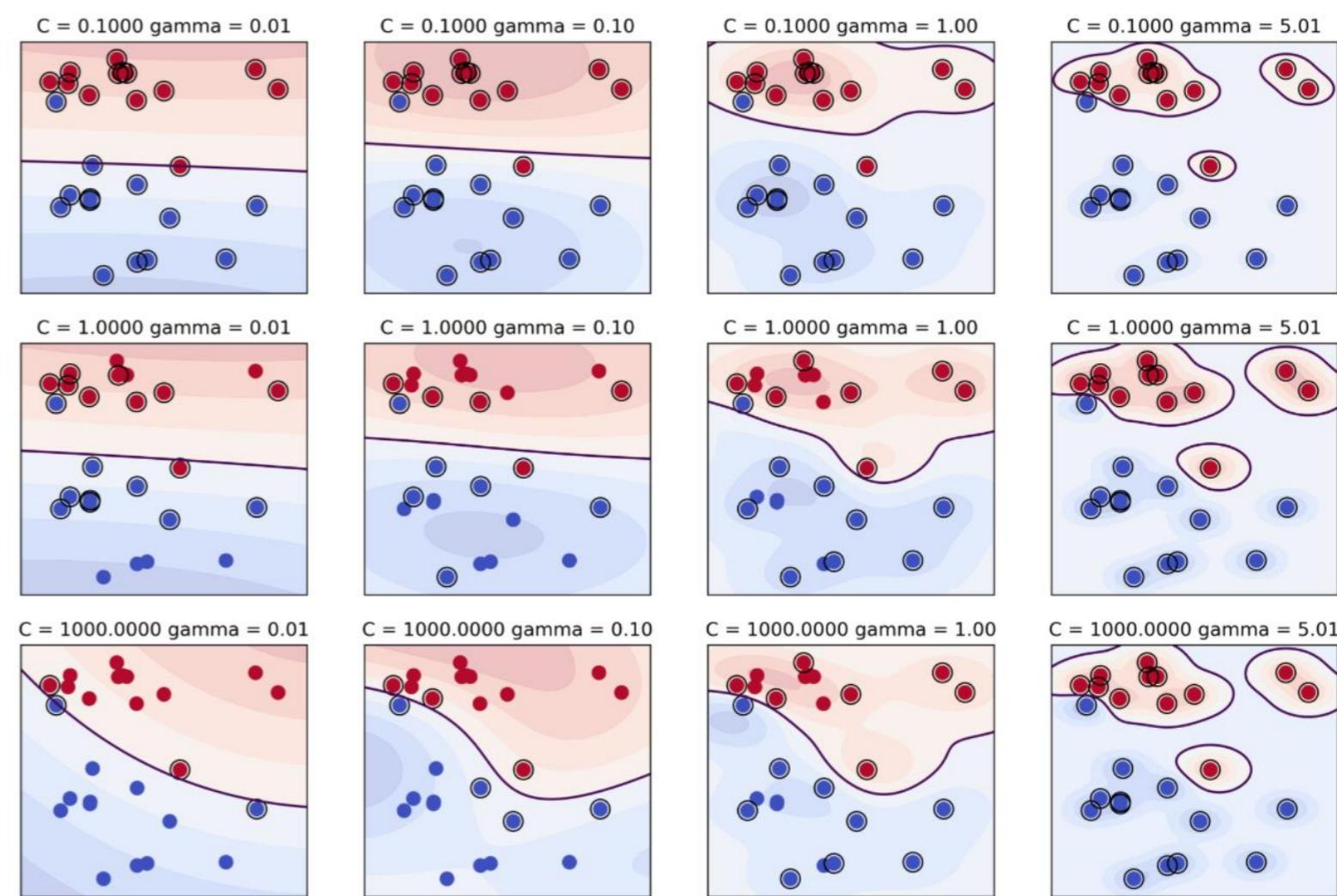
**Over-fitting**  
(forcefitting--too  
good to be true)

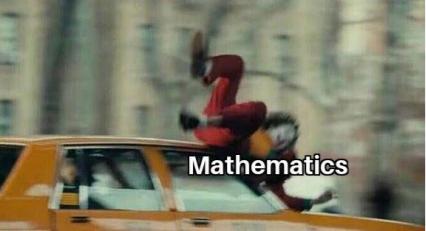
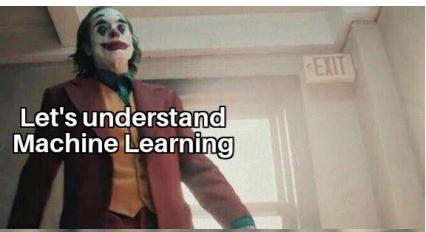


Data with kernel



SVM - Gamma

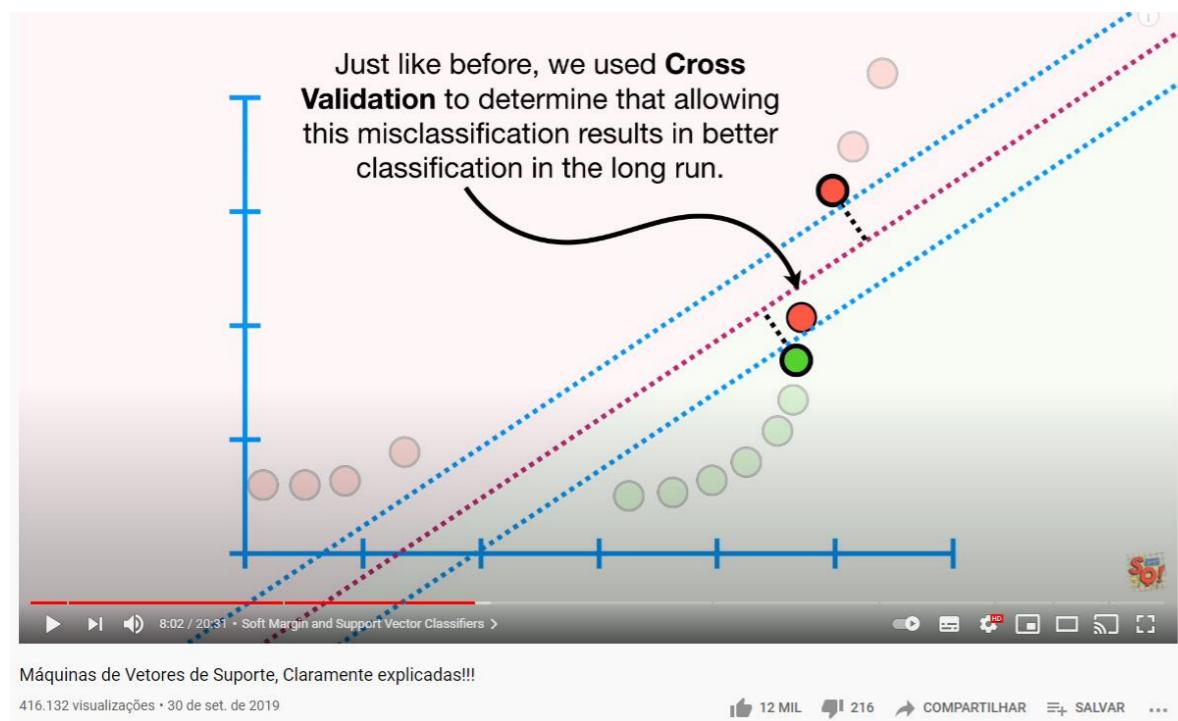




# SUPPORT VECTOR MACHINES (SVM)

*Want to know more??*

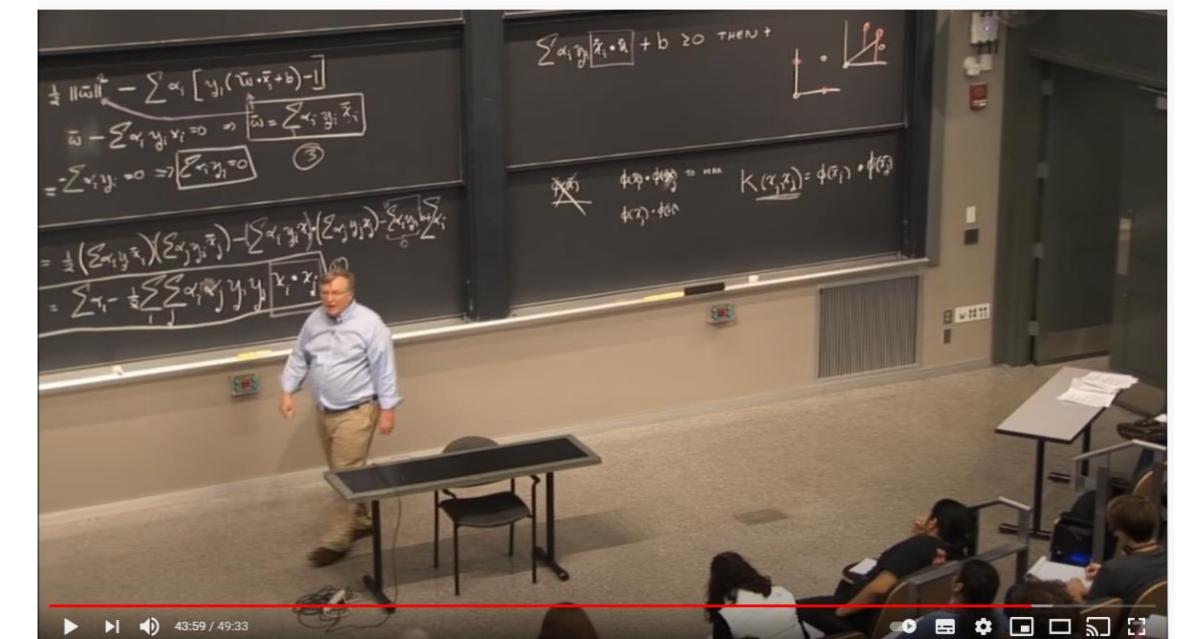
**Easy mode**



**StatQuest with Josh Starmer**

<https://www.youtube.com/watch?v=efR1C6CvhmE>

**Math and statistics required!!**



16. Learning: Support Vector Machines

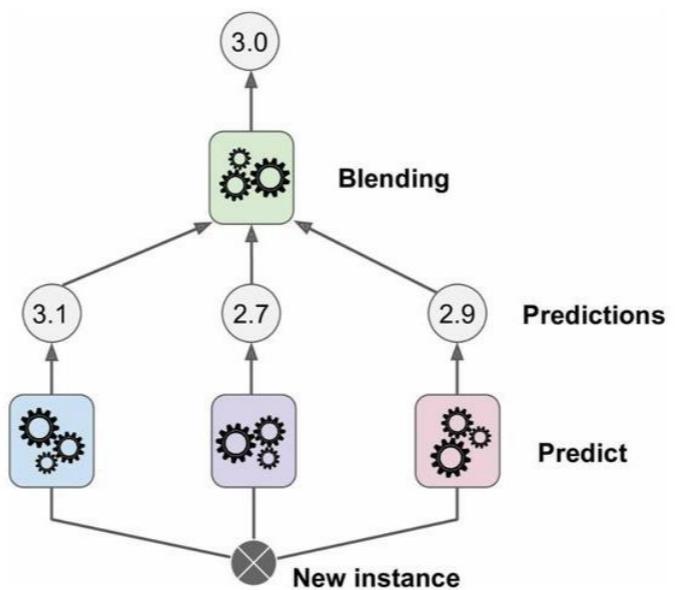
1.223.895 visualizações • 10 de jan. de 2014

13 MIL 151 COMPARILHAR SALVAR ...

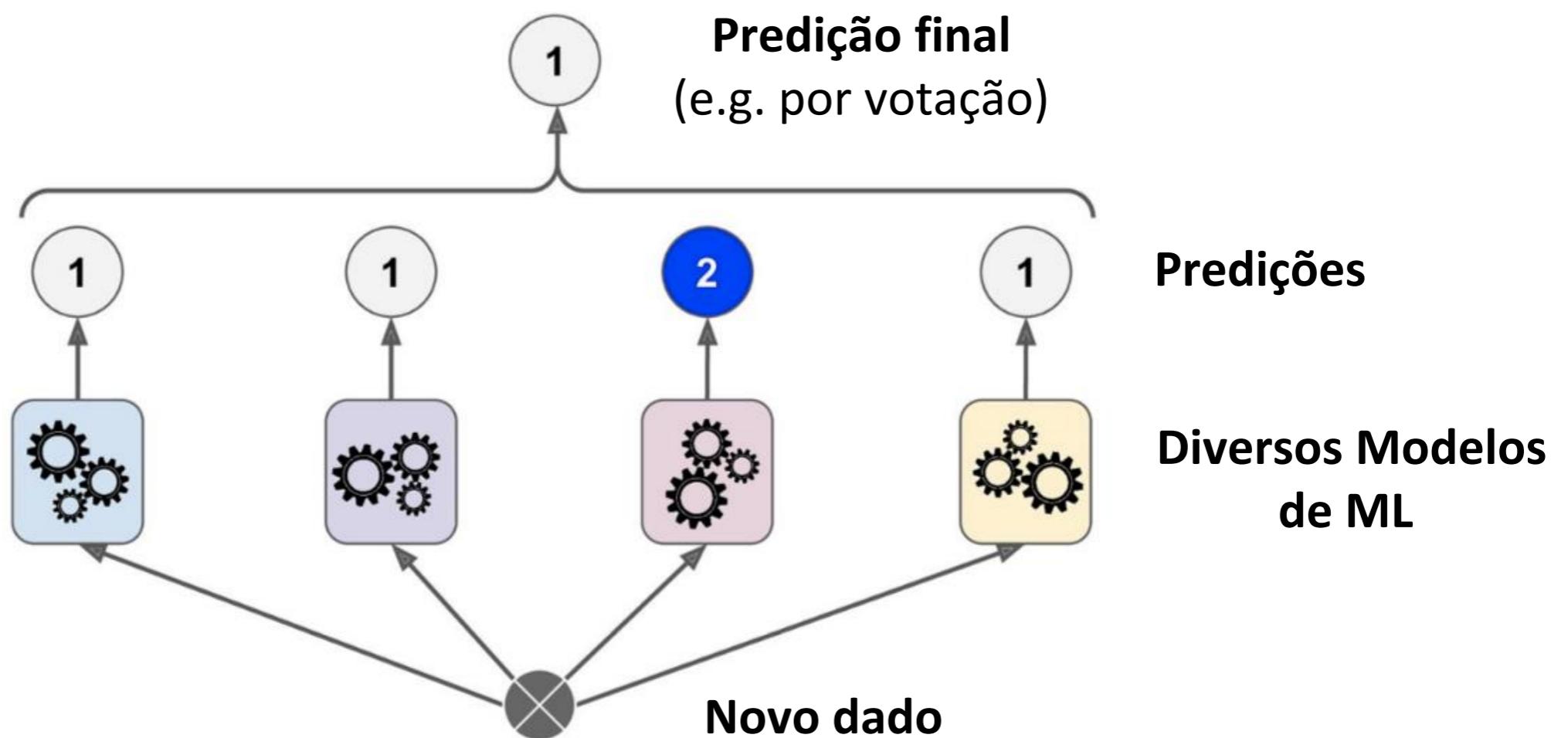
**MIT OpenCourseWare**

[https://youtu.be/\\_PwhiWxHK8o](https://youtu.be/_PwhiWxHK8o)

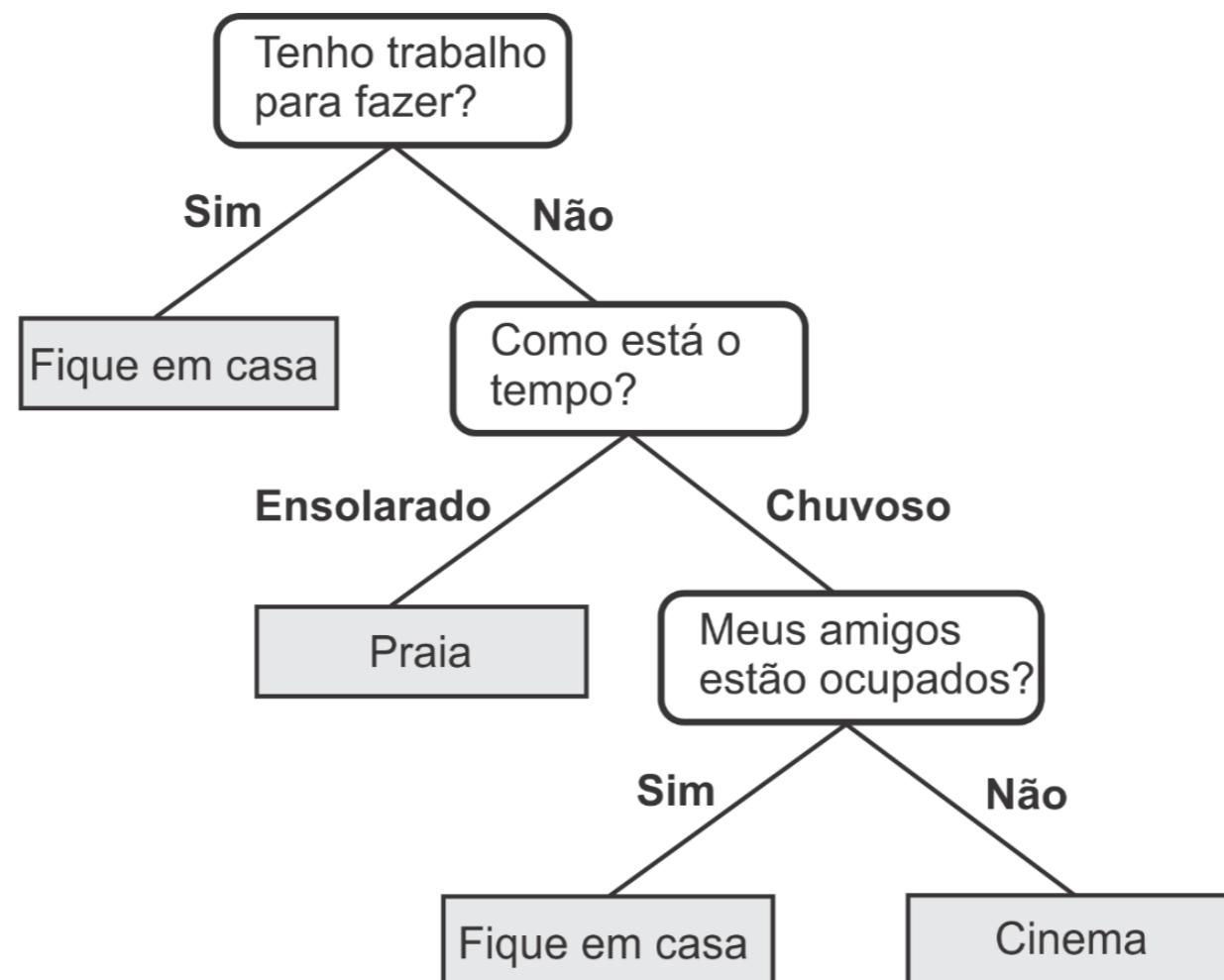
# Ensemble Methods



# Ensemble Methods



# Árvores de Decisão

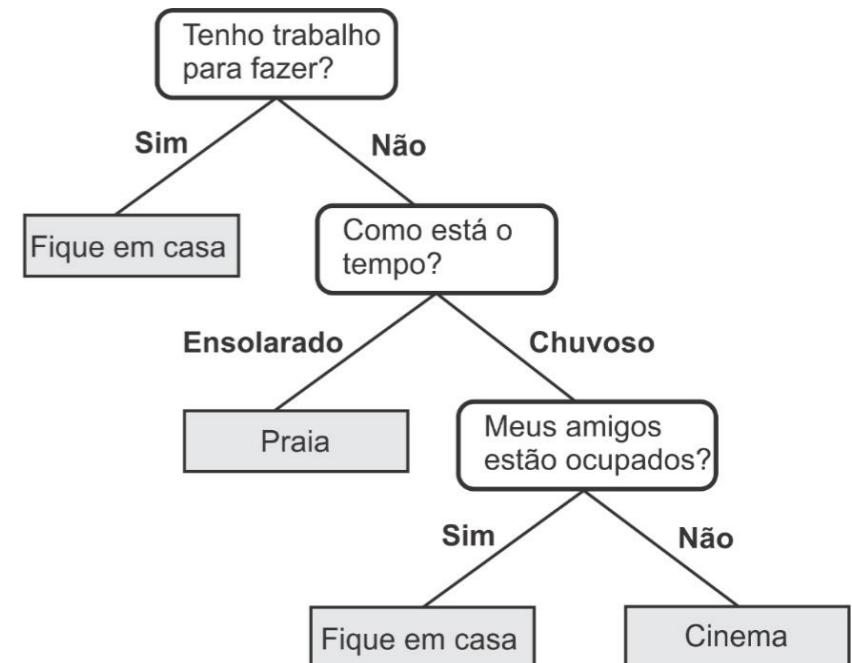


# Random Forest



# Random Forest

- **Random Forest (RF)** é um algoritmo de *machine learning* para classificação e regressão de dados, que combina múltiplas **árvores de decisão** para realizar previsões.
- RF é robusto a ruídos na base de dados de treinamento
- RF é rápido e efetivo para processar grandes bases de dados



# Random Forest

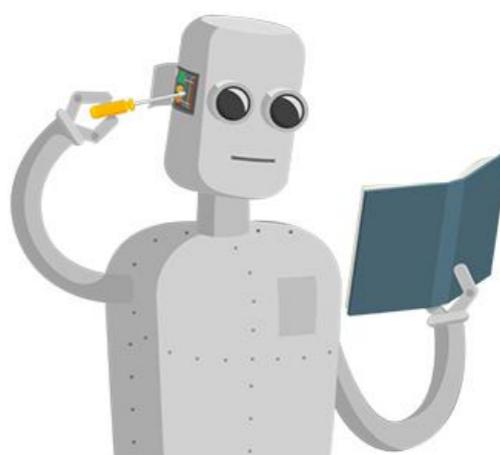
## Exemplo

### Dados de Treinamento

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,11	5,14	1,62	2,70	306,83	A4_mu_c
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

*Features*      *Target*

Vamos  
aprender!



# Random Forest

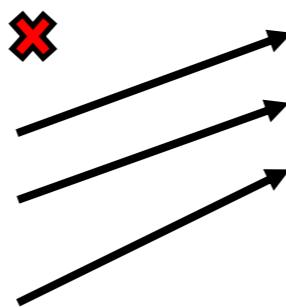
1º) Selecionar aleatoriamente 2/3 dos dados (ou *instances*) colocar em uma “bag”

**Dados de Treinamento**

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,11	5,14	1,62	2,70	306,83	A4_mu_c
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

**Bootstrapped Data (~2/3)**

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c
**	**	**	**	**	**	**



# Random Forest

2º) Repetir instâncias aleatórias dentro da “bag” para ficar do mesmo tamanho do dado original

Bootstrapped Data (~2/3)



Bootstrapped Data (~2/3) + (~1/3) replacement

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c
<b>0,18</b>	<b>0,13</b>	<b>5,42</b>	<b>2,20</b>	<b>3,19</b>	<b>287,05</b>	<b>A4_mu_c</b>

Dado repetido (*replacement*)

# Random Forest

3º) Selecionar *features* aleatórias (mtry)

Bootstrapped Data (~2/3) + (~1/3) replacement

GT (nT/m )	K (%)	eTh (ppm )	eU (ppm )	CT ( $\mu$ R/h)	Elevaçã o (m)	Litologi a
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_ c
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_ c

$mtry \leq \sqrt{numero\ de\ features}$

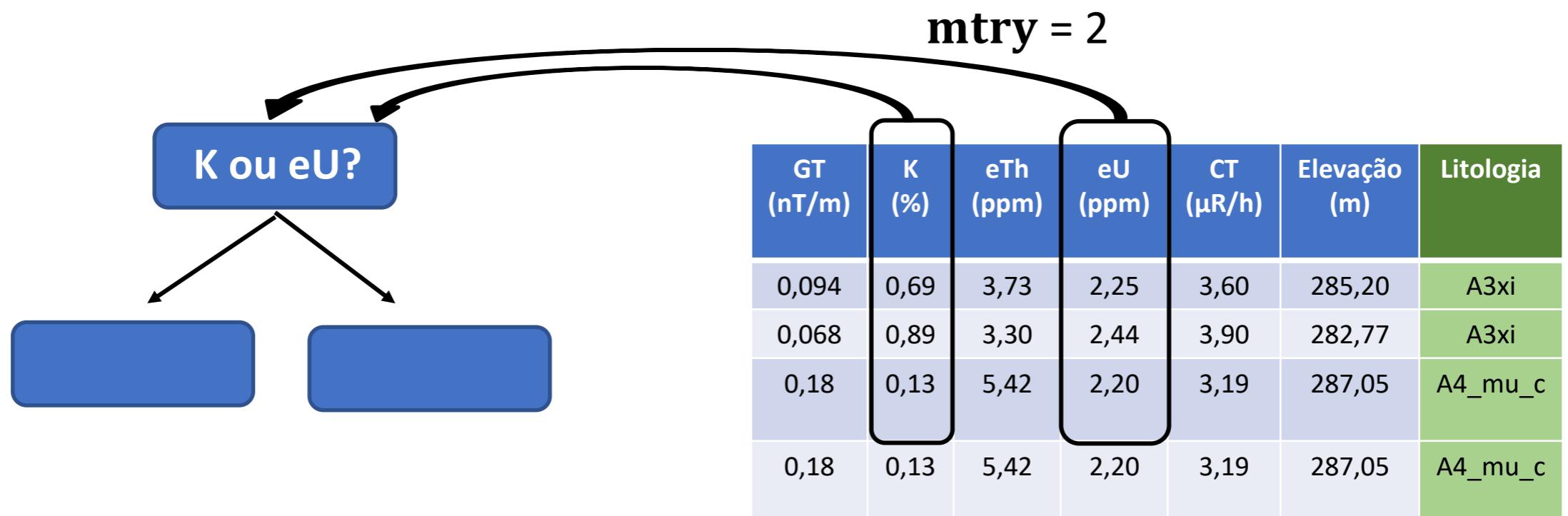
$mtry \leq \sqrt{6}$

$mtry \leq \sim 2.45$

Seleciona aleatoriamente 2 variáveis!

# Random Forest

3º) Selecionar *features* aleatórias (mtry)

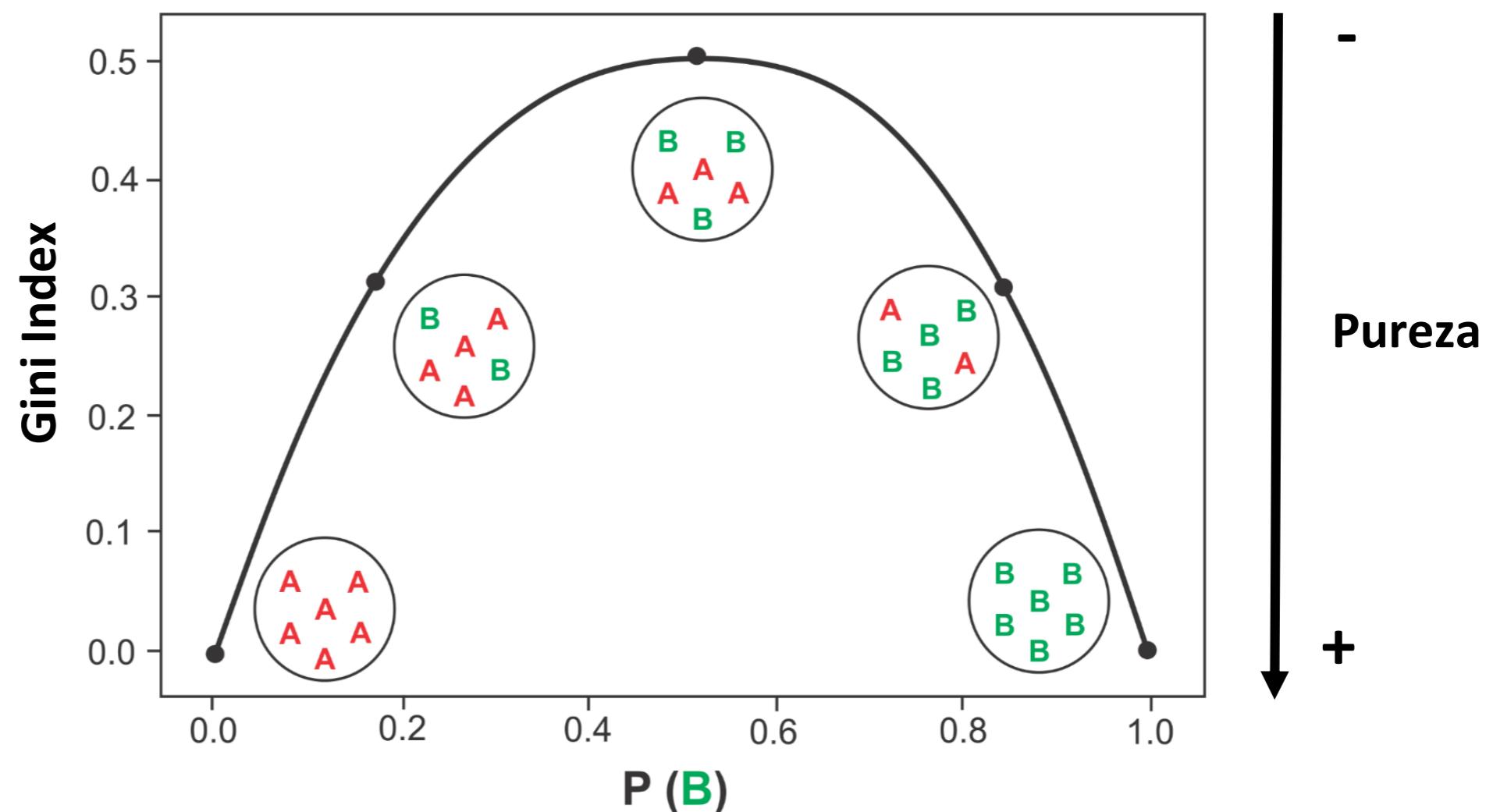


1) Qual *feature* separa melhor os dados? Potássio (K) ou Urânio (eU)?

2) Qual o valor da feature escolhida que melhor separa os dados?

**Pureza!**

# Gini Index



# Gini Index

**Potássio (K)**

Litologia	Potássio (K)
A	0
A	0.1
A	0.2
B	0.3
B	0.4
B	0.5
B	0.6
B	0.7

$$Gini = 0.2$$

**Urânio (eU)**

Litologia	Urânio (eU)
A	0
B	2
A	4
B	5
B	7
B	9
B	11
B	12

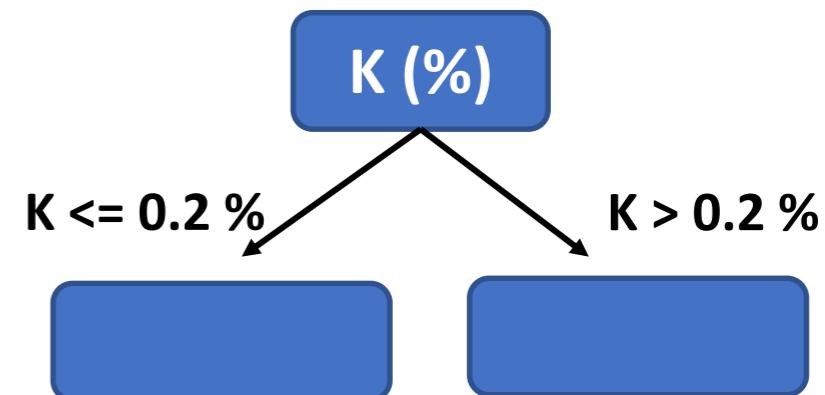
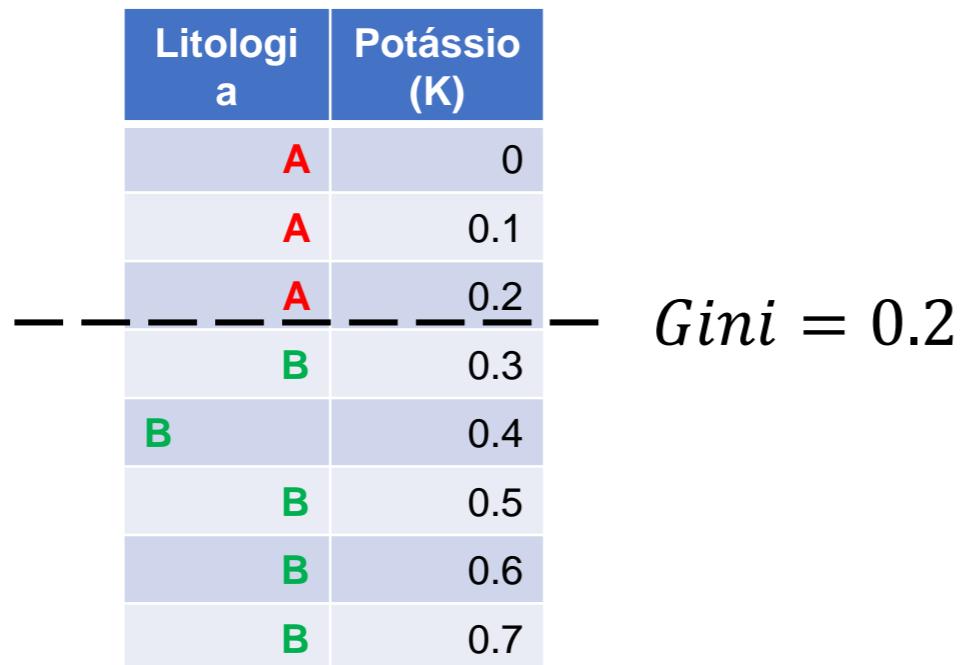
$$Gini = 0.3$$



Potássio (K) é a *feature* que melhor separa os dados!  
E o melhor valor de separação é 0.2!

# Gini Index

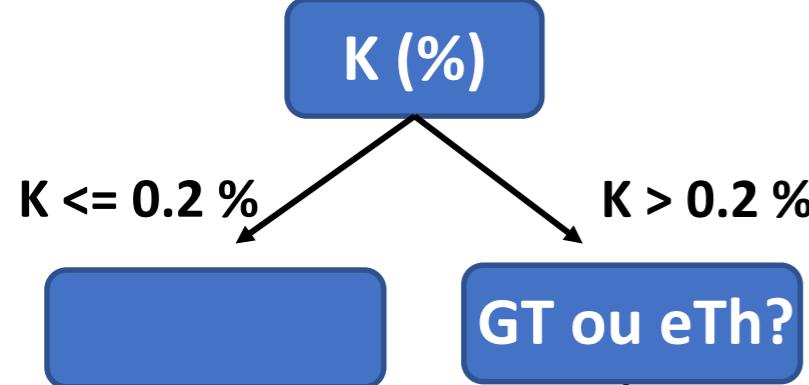
## Potássio (K)



Potássio (K) é a *feature* que melhor separa os dados!  
E o melhor valor de separação é 0.2!

# Random Forest

4º) Construir a árvore de decisão minimizando a impureza nos nós subsequentes

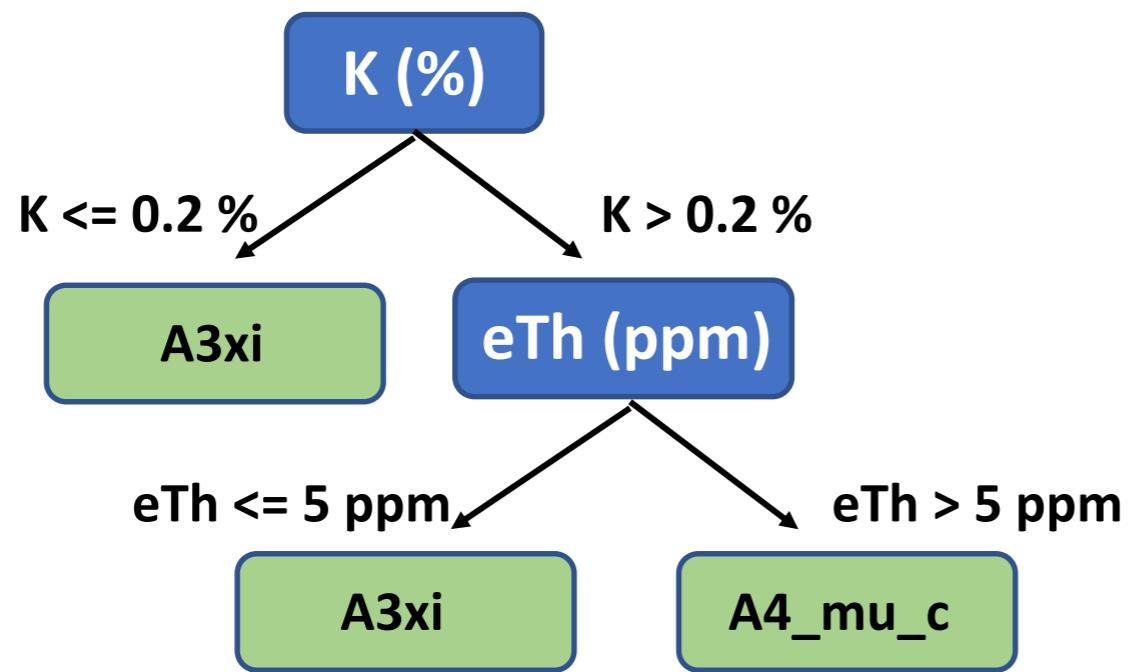


Qual variável  
separa melhor?  
GT ou eTh?

Conjunto de dados com  $K > 0.2 \%$

GT (nT/m )	K (%)	eTh (ppm )	eU (ppm )	CT ( $\mu$ R/h)	Elevaçã o (m)	Litologi a
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,23	5,42	2,20	3,19	287,05	A4_mu_c
0,18	0,23	5,42	2,20	3,19	287,05	A4_mu_c

# Random Forest



1ª Árvore de decisão!

# Random Forest

4.1º) Repetir esse procedimento para diferentes conjuntos de dados aleatórios (*Bootstrapped Data* )

**Dados de Treinamento**

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,11	5,14	1,62	2,70	306,83	A4_mu_c
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

**Bootstrapped Data (~2/3) + (1/3) replacement**

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c
<b>0,18</b>	<b>0,13</b>	<b>5,42</b>	<b>2,20</b>	<b>3,19</b>	<b>287,05</b>	<b>A4_mu_c</b>

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,11	5,14	1,62	2,70	306,83	A4_mu_c
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,11	5,14	1,62	2,70	306,83	A4_mu_c
<b>0,219</b>	<b>0,11</b>	<b>5,14</b>	<b>1,62</b>	<b>2,70</b>	<b>306,83</b>	<b>A4_mu_c</b>
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi

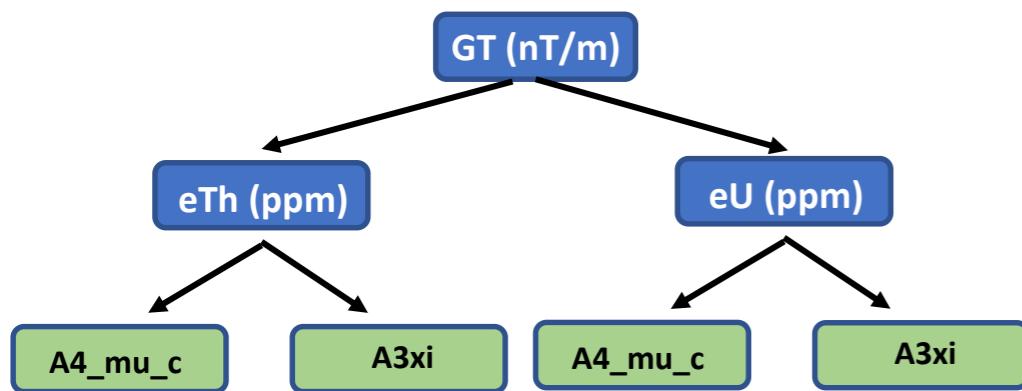
\*Dado repetido (replacement)

1º

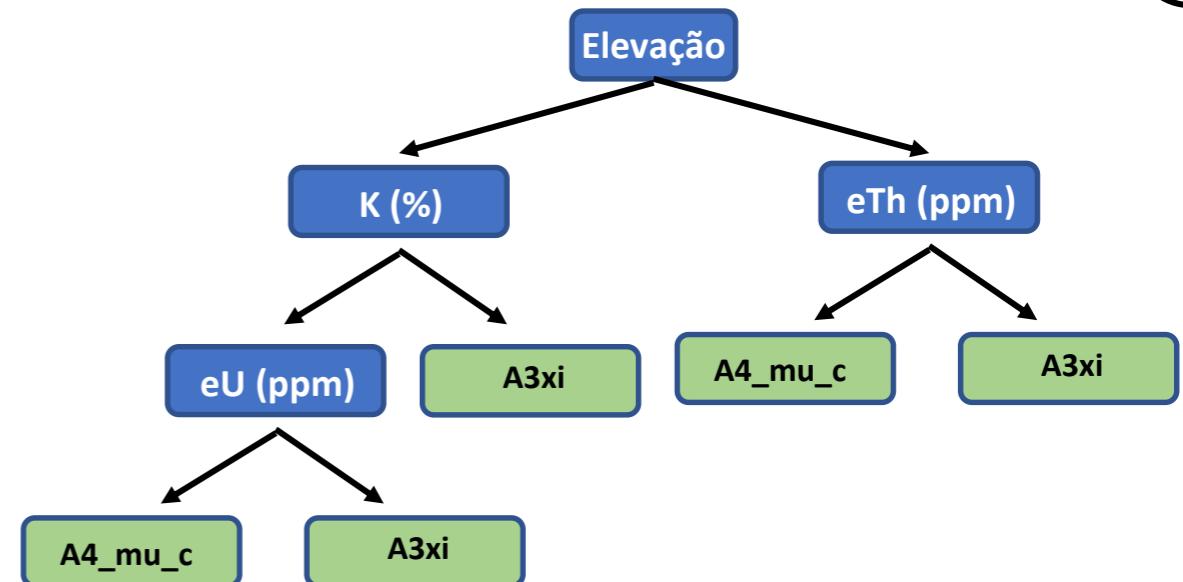
2º

# Random Forest

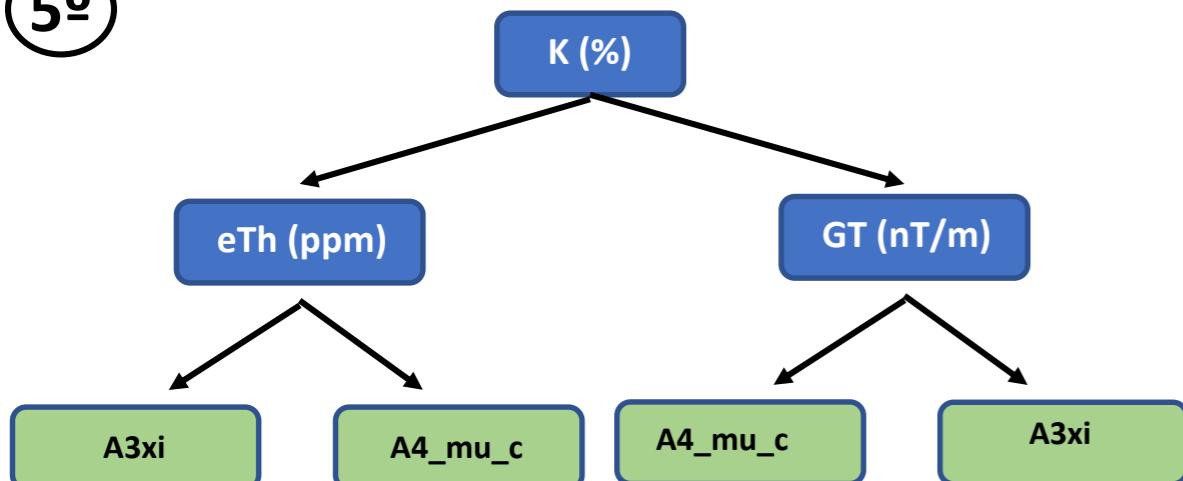
3º



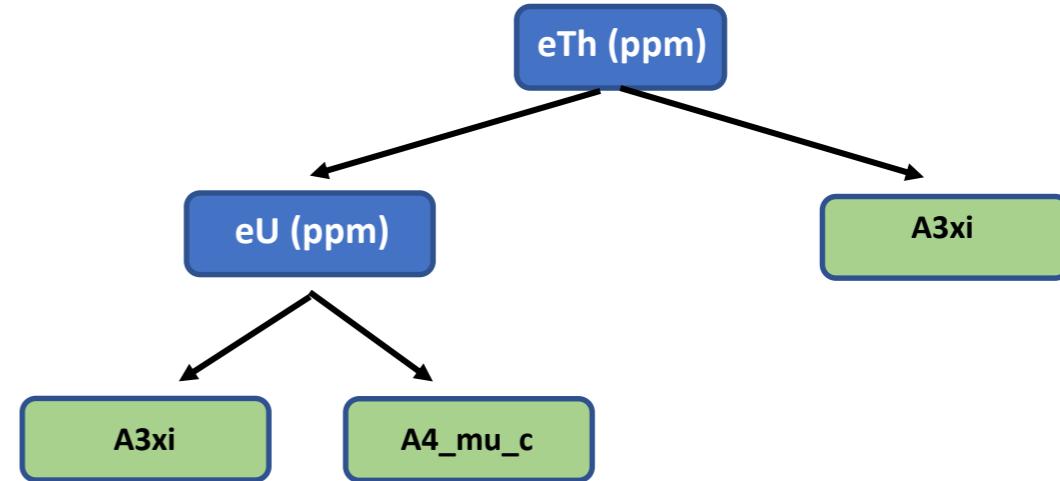
4º



5º



6º

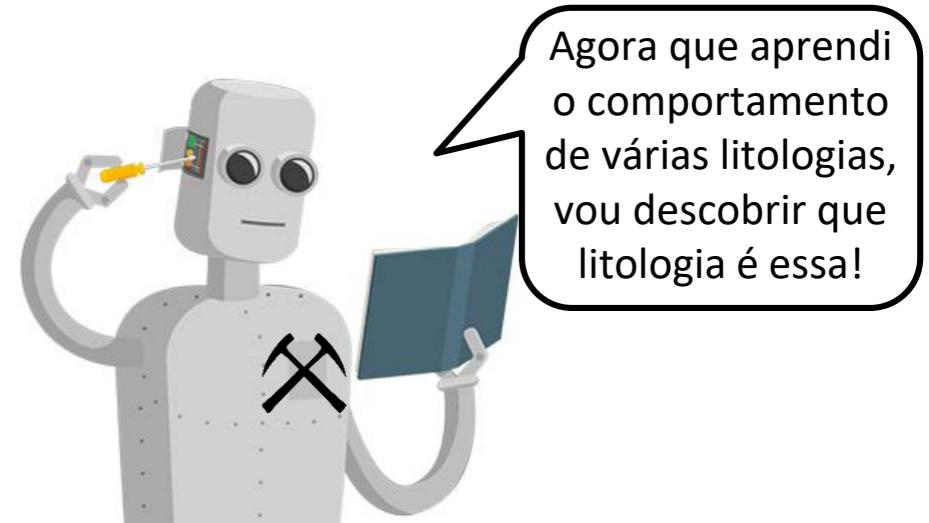


...

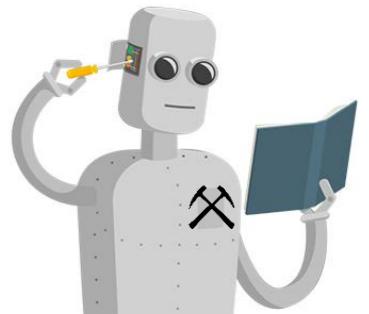
# Random Forest

## Litologia desconhecida

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,166	0,25	4,79	2,34	3,56	285,27	???



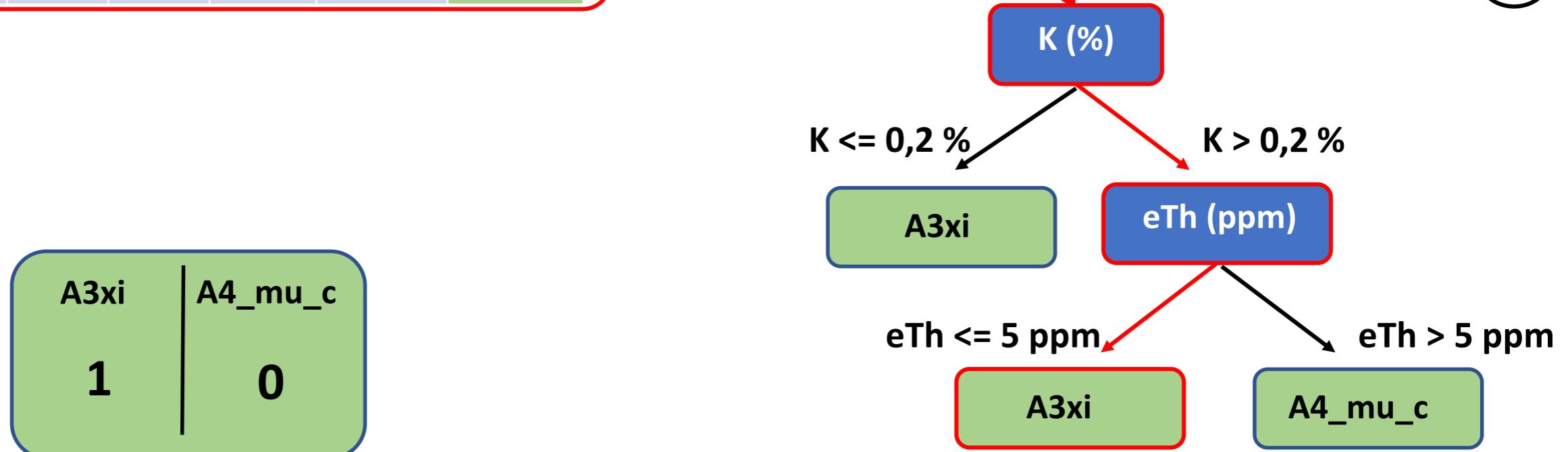
# Random Forest



**Litologia desconhecida**

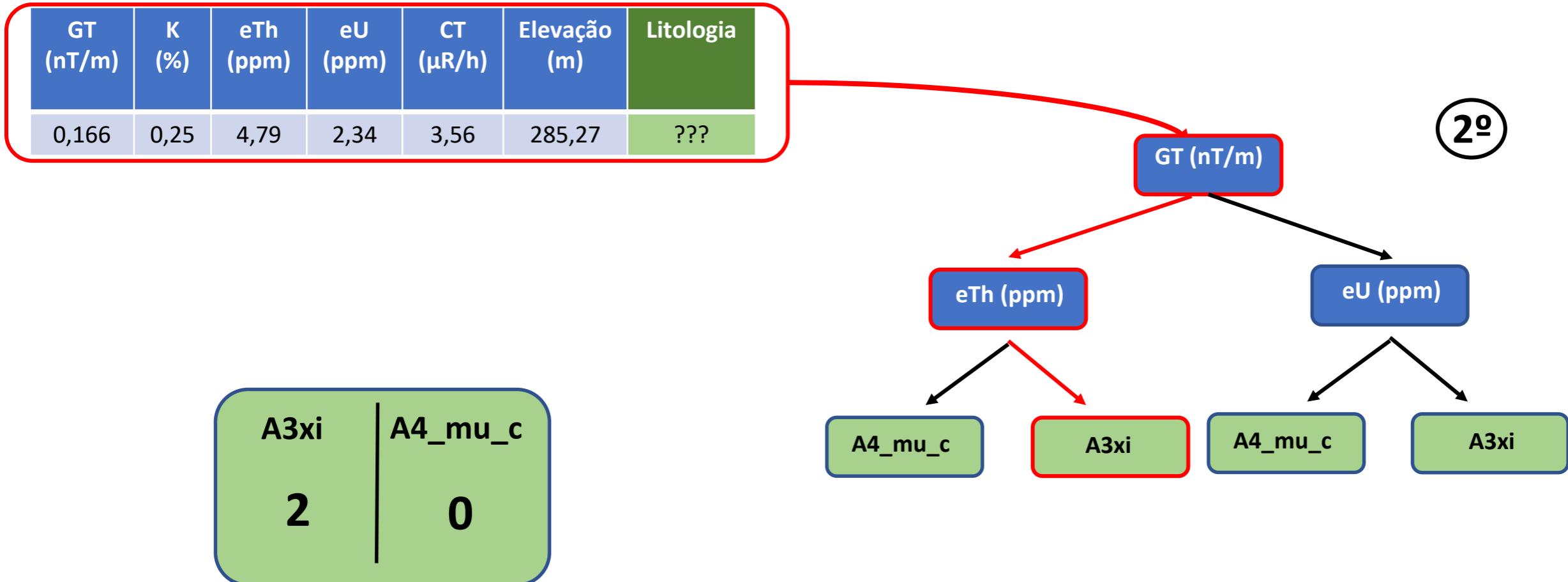
GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,166	0,25	4,79	2,34	3,56	285,27	???

1º



# Random Forest

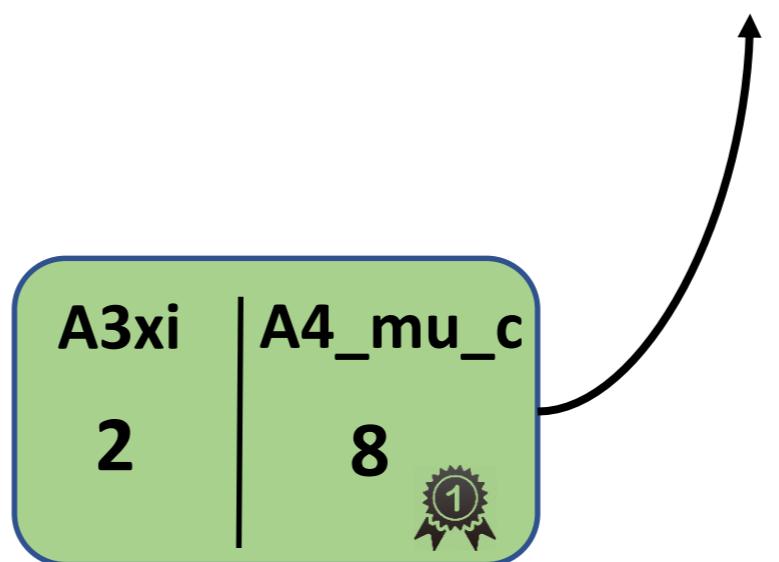
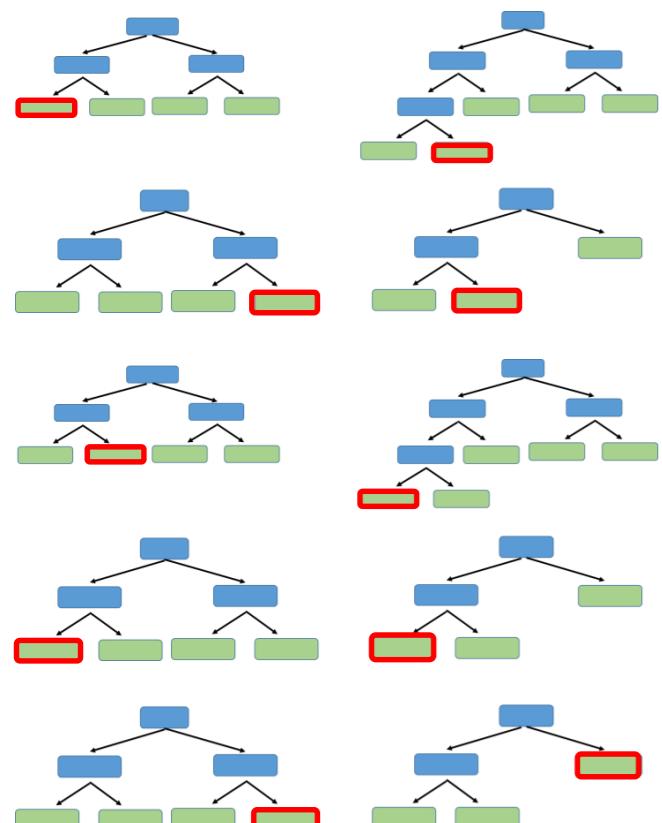
## Litologia desconhecida



*And so on...*

# Random Forest

GT (nT/m )	K (%)	eTh (ppm )	eU (ppm )	CT ( $\mu$ R/h)	Elevaçã o (m)	Litologi a
0,166	0,25	4,79	2,34	3,56	285,27	A4_mu_c



*Mas como definir a quantidade de árvores de decisão?*

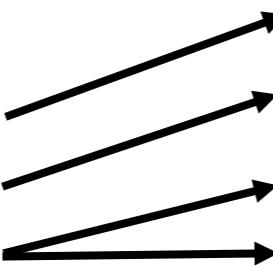
# Random Forest

1º

Dados de Treinamento

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,21	4,95	1,62	2,70	306,83	A4_mu_c
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

✗



Bootstrapped Data (~2/3)

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,094	0,69	3,73	2,25	3,60	285,20	A3xi
0,068	0,89	3,30	2,44	3,90	282,77	A3xi
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c
0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

“Out-Of-Bag Data”

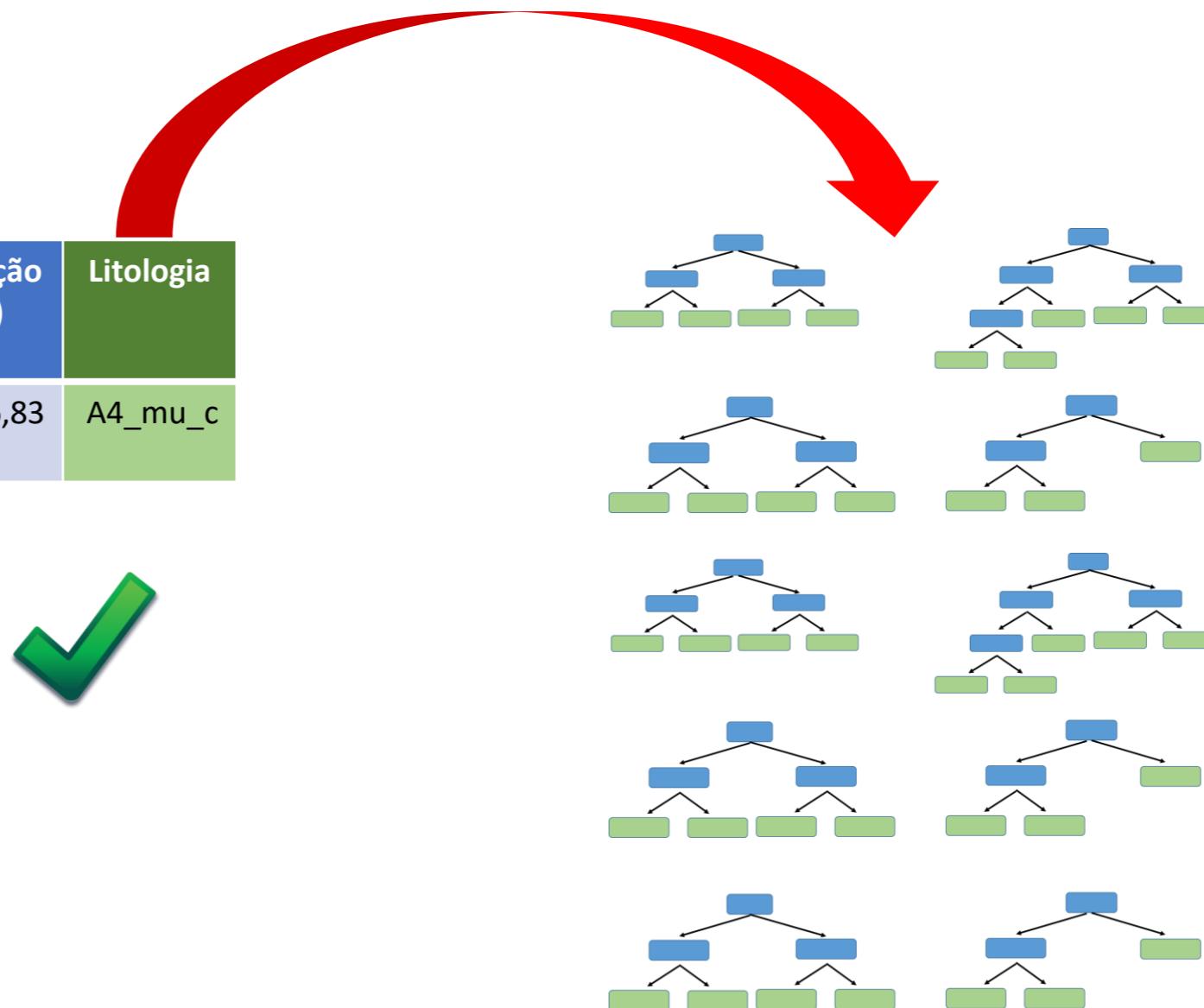
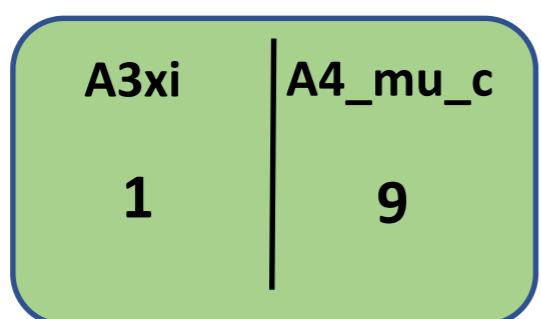
GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,21	4,95	1,62	2,70	306,83	A4_mu_c



# Random Forest

“Out-Of-Bag Data”

GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
0,219	0,21	4,95	1,62	2,70	306,83	A4_mu_c



# Random Forest

OOB 1	GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
	0,219	0,21	5,14	1,62	2,70	306,83	A4_mu_c

A3xi 1	A4_mu_c 9
-----------	--------------



OOB 2	GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
	0,094	0,69	3,73	2,25	3,60	285,20	A3xi

A3xi 8	A4_mu_c 2
-----------	--------------



“Out-of-Bag Error”

OOB 3	GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
	0,068	0,89	3,30	2,44	3,90	282,77	A3xi

A3xi 8	A4_mu_c 2
-----------	--------------

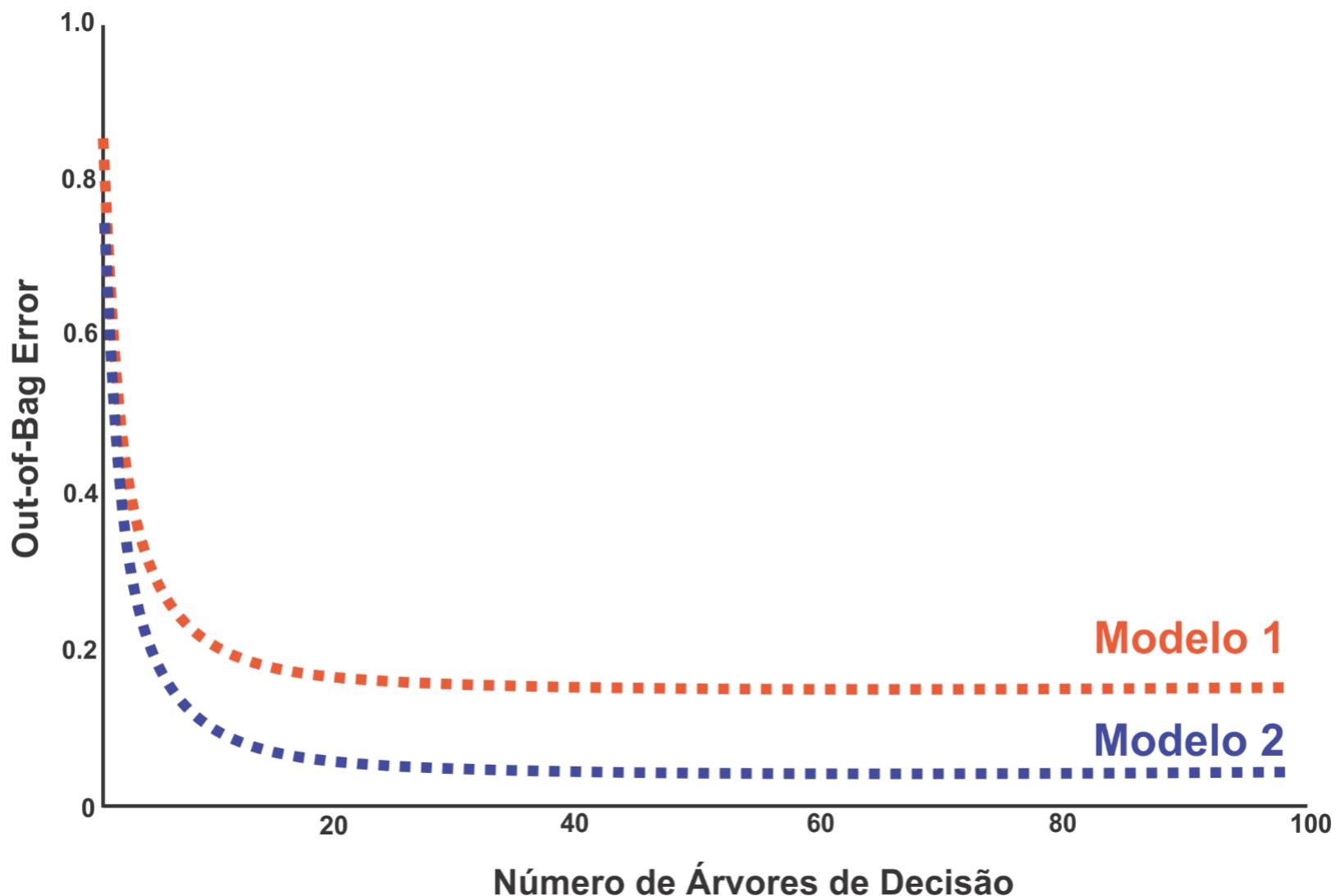


OOB 4	GT (nT/m)	K (%)	eTh (ppm)	eU (ppm)	CT ( $\mu$ R/h)	Elevação (m)	Litologia
	0,18	0,13	5,42	2,20	3,19	287,05	A4_mu_c

A3xi 6	A4_mu_c 4
-----------	--------------



# Random Forest



# Random Forest nas Geociências

**Carranza E J M, & Laborte A G. 2015a.** Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of random forests algorithm. *Ore Geol Rev*, 71: 777–787

**Hariharan, S. et al. (2017).** Random forest-based prospectivity modelling of greenfield terrains using sparse deposit data: An example from the Tanami Region, Western Australia. *Natural Resources Research*, 26(4), 489-507.

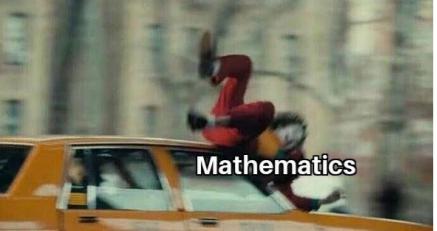
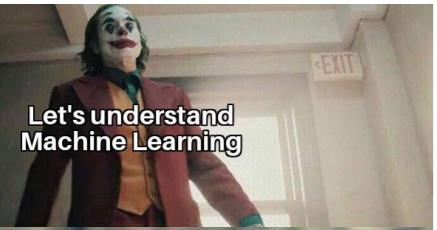
**McKay, G., & Harris, J. R. (2016).** Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, 25(2), 125-143.

**Zhang, Z. et al. 2016.** A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, 59(3), 556-572.

**Ford, A. (2020).** Practical implementation of random forest-based mineral potential mapping for porphyry Cu–Au mineralization in the Eastern Lachlan Orogen, NSW, Australia. *Natural Resources Research*, 29(1), 267-283.

**Carranza E J M, & Laborte A G. 2015b.** Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Comput Geosci*, 74: 60–70

**Rodríguez-Galiano et al. 2014.** Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain. *Int J Geogr Inf Sci*, 28: 1336–1354



# Random Forest

*Want to know more??*

**Easy mode**

So we take the data and run it down the first tree that we made...

Ches Pain Good Blood Circ. Blocked Arteries Weight Heart Disease

Yes	No	No	168	Heart Disease
-----	----	----	-----	---------------

The video shows a decision tree with four nodes. The first node (red box) splits into two branches. The left branch leads to a green node, which then splits into two green leaf nodes. The right branch leads to another green node, which then splits into two green leaf nodes. A red arrow points from the text 'So we take the data and run it down the first tree that we made...' to the first red box on the tree. A red oval highlights the 'Heart Disease' column in the data table, which shows '168'.

4:52 / 9:53 • Classifying a new sample with a Random Forest >

StatQuest: Random Forests Parte 1 - Construindo, Usando e Avaliando

496.638 visualizações • 5 de fev. de 2018

8,9 MIL 118 COMPARTELHAR SALVAR ...

**StatQuest with Josh Starmer**

[https://www.youtube.com/watch?v=J4Wdy0Wc\\_xQ](https://www.youtube.com/watch?v=J4Wdy0Wc_xQ)

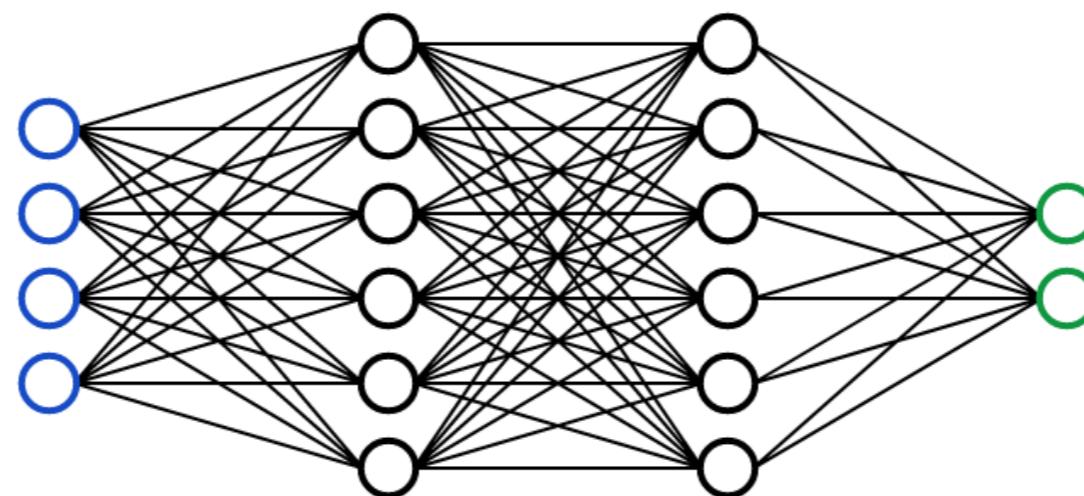
**Math and statistics required!!**



**Kilian Weinberger**

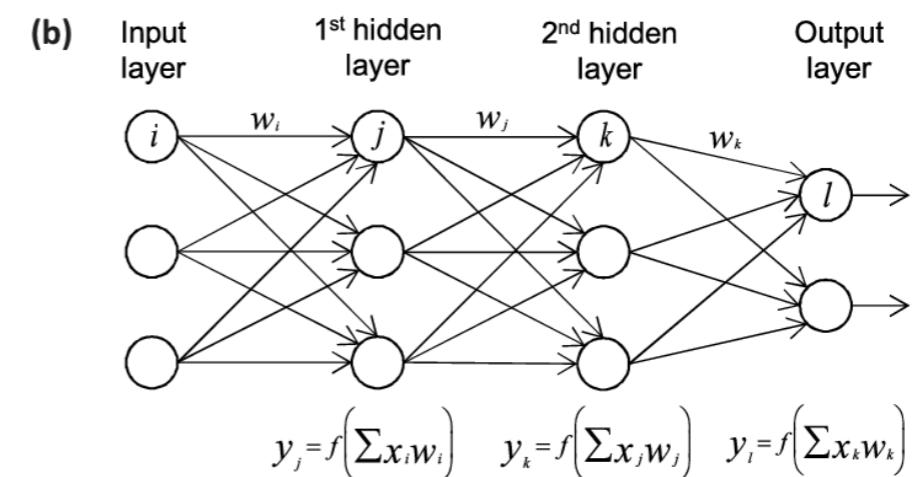
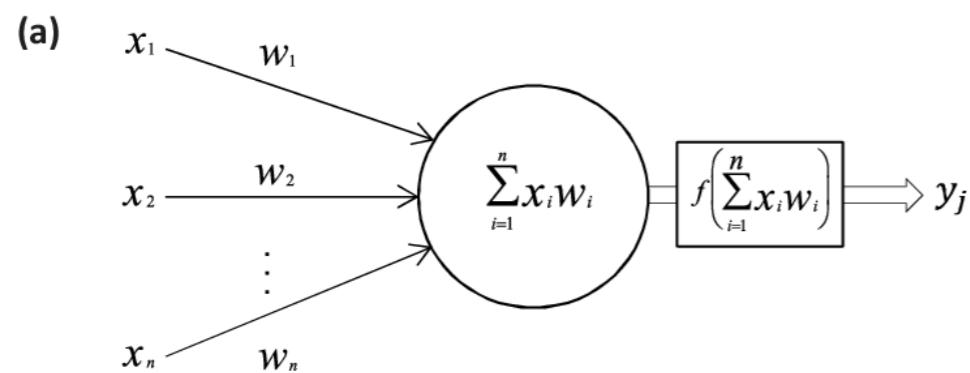
<https://www.youtube.com/watch?v=4EOCQJgqAOY>

# Neural Networks



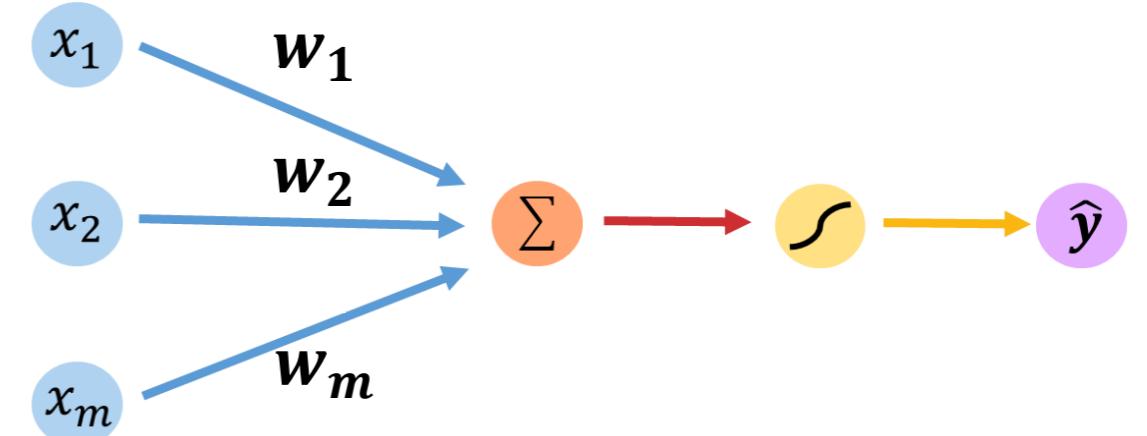
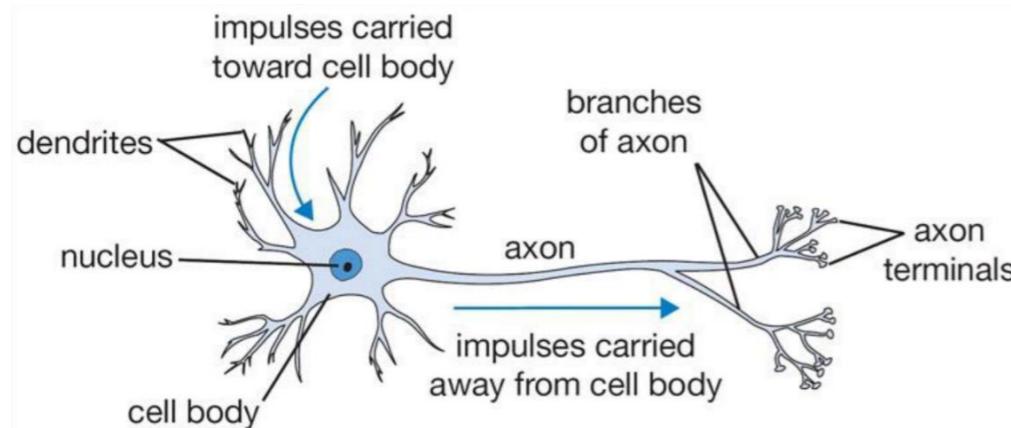
# Neural Network

- **Neural Network (NN)** é um dos algoritmos mais famosos em Machine Learning
- No entanto, **NN** é comumente associado a algoritmos *black box*, uma vez pode ser complicado entender como é construída sua inteligência



# Neural Network

## Neurons



Inputs

Weights

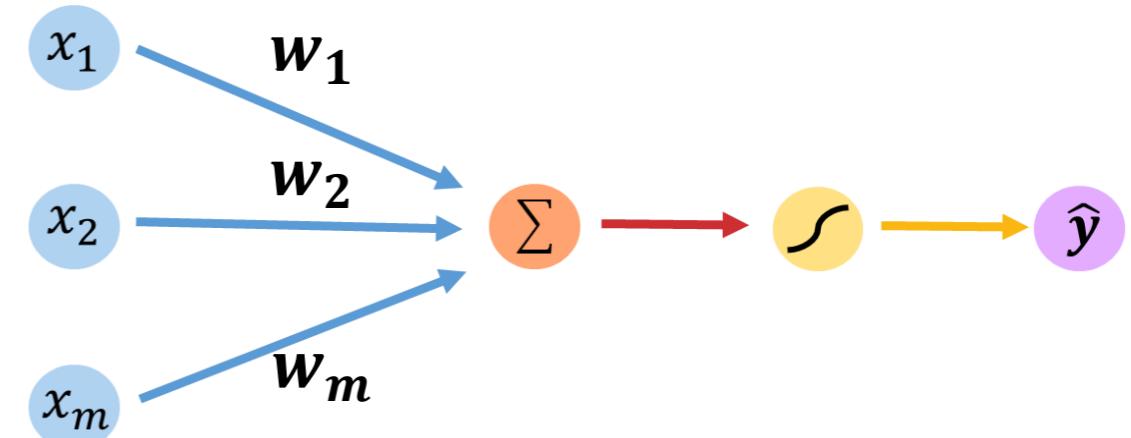
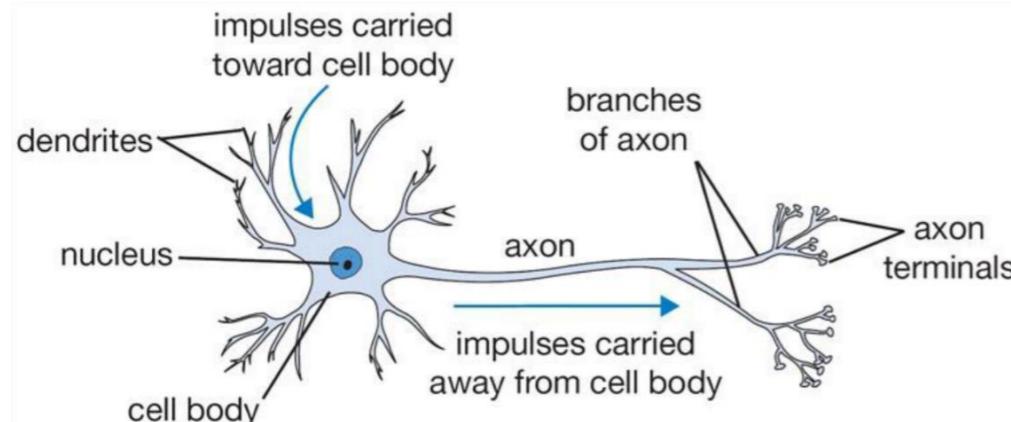
Sum

Non-Linearity

Output

# Neural Network

## Neurons



Inputs      Weights      Sum      Non-Linearity      Output

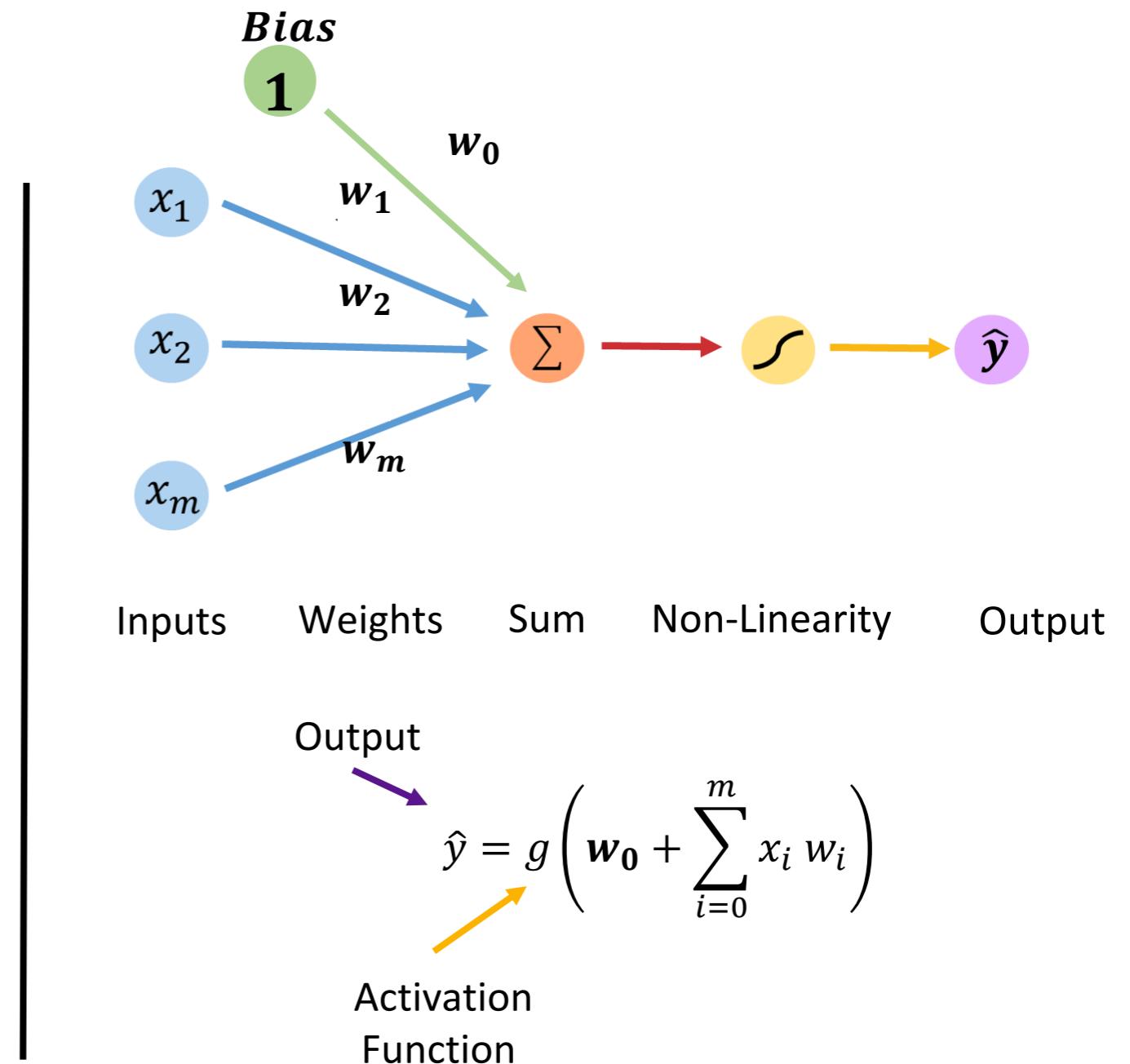
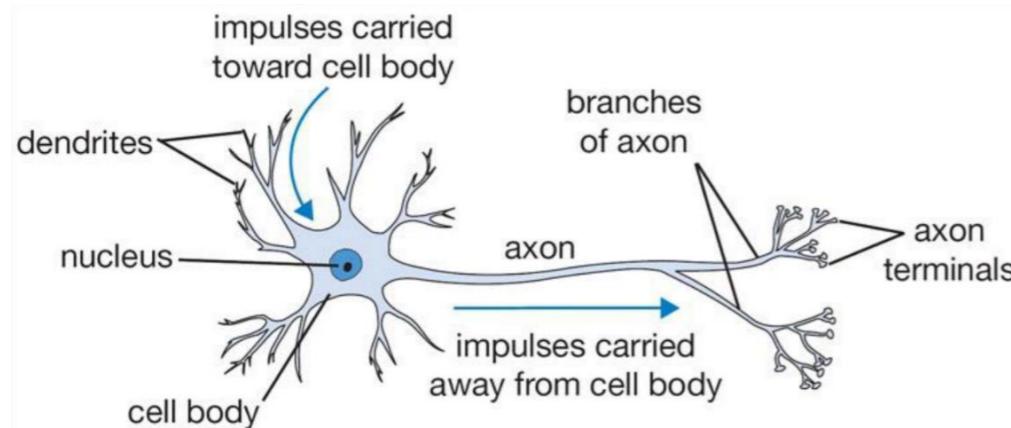
Output

$$\hat{y} = g \left( \sum_{i=0}^m x_i w_i \right)$$

Activation  
Function

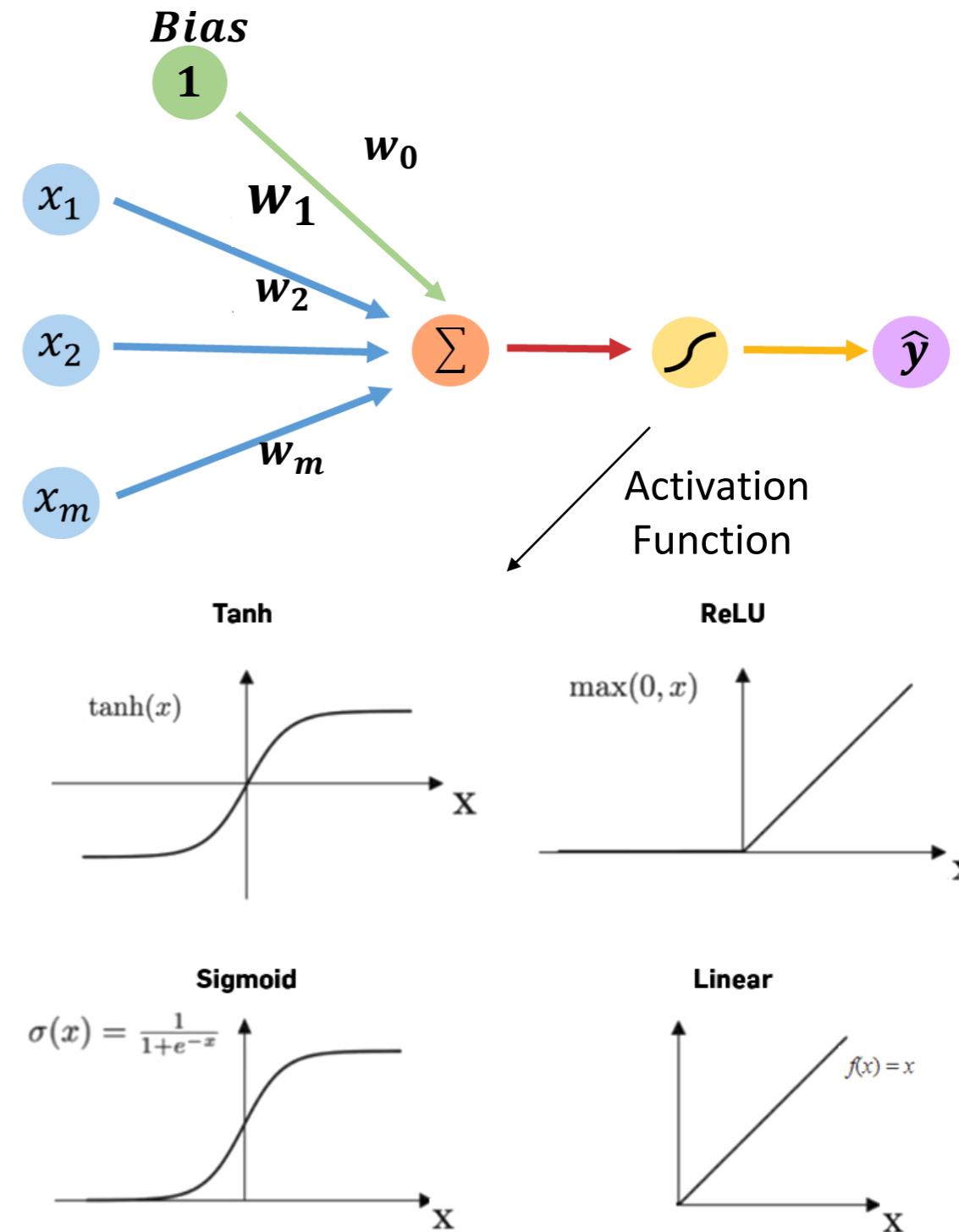
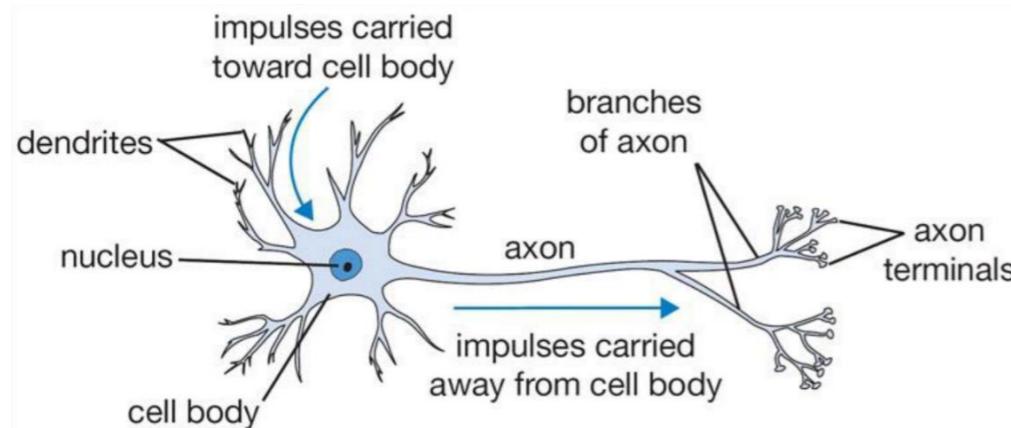
# Neural Network

## Neurons

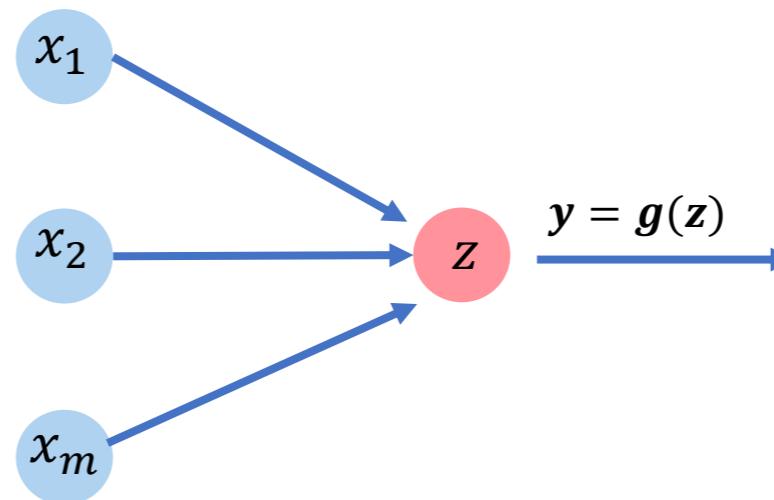
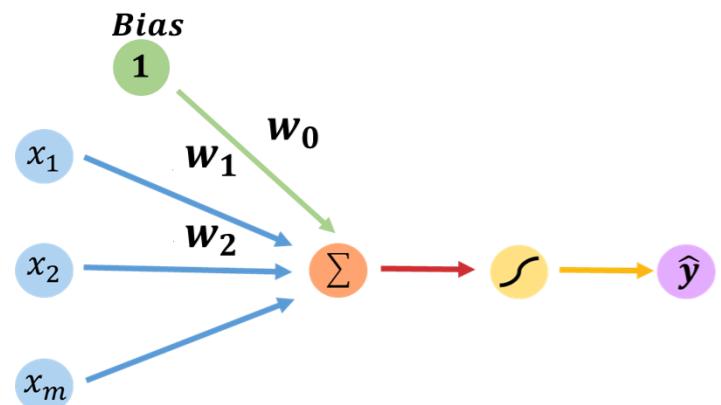


# Neural Network

## Neurons

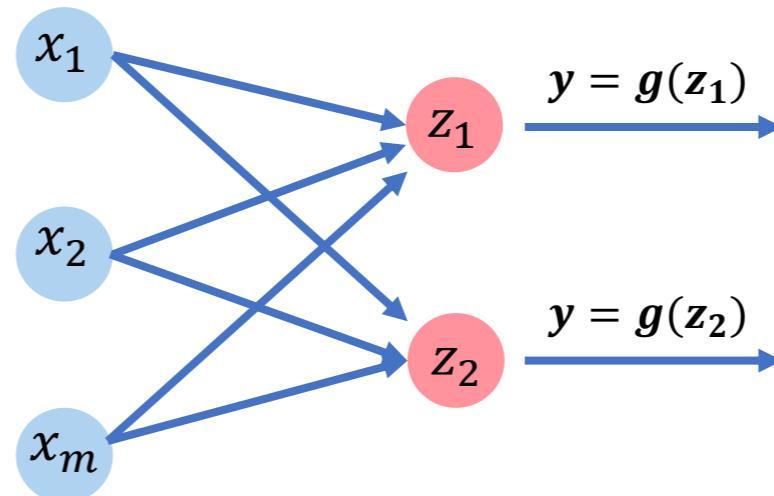
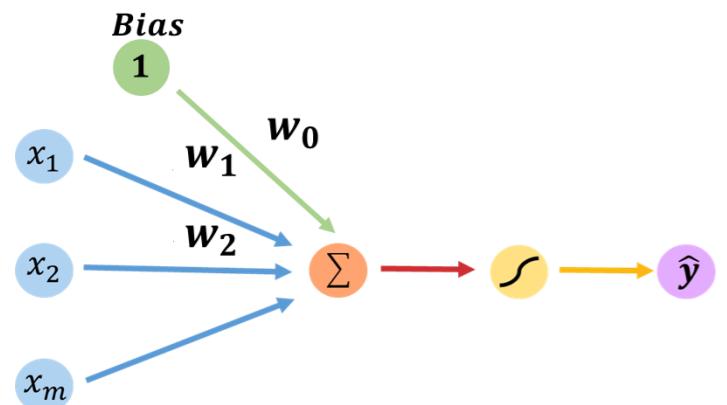


# Neural Network



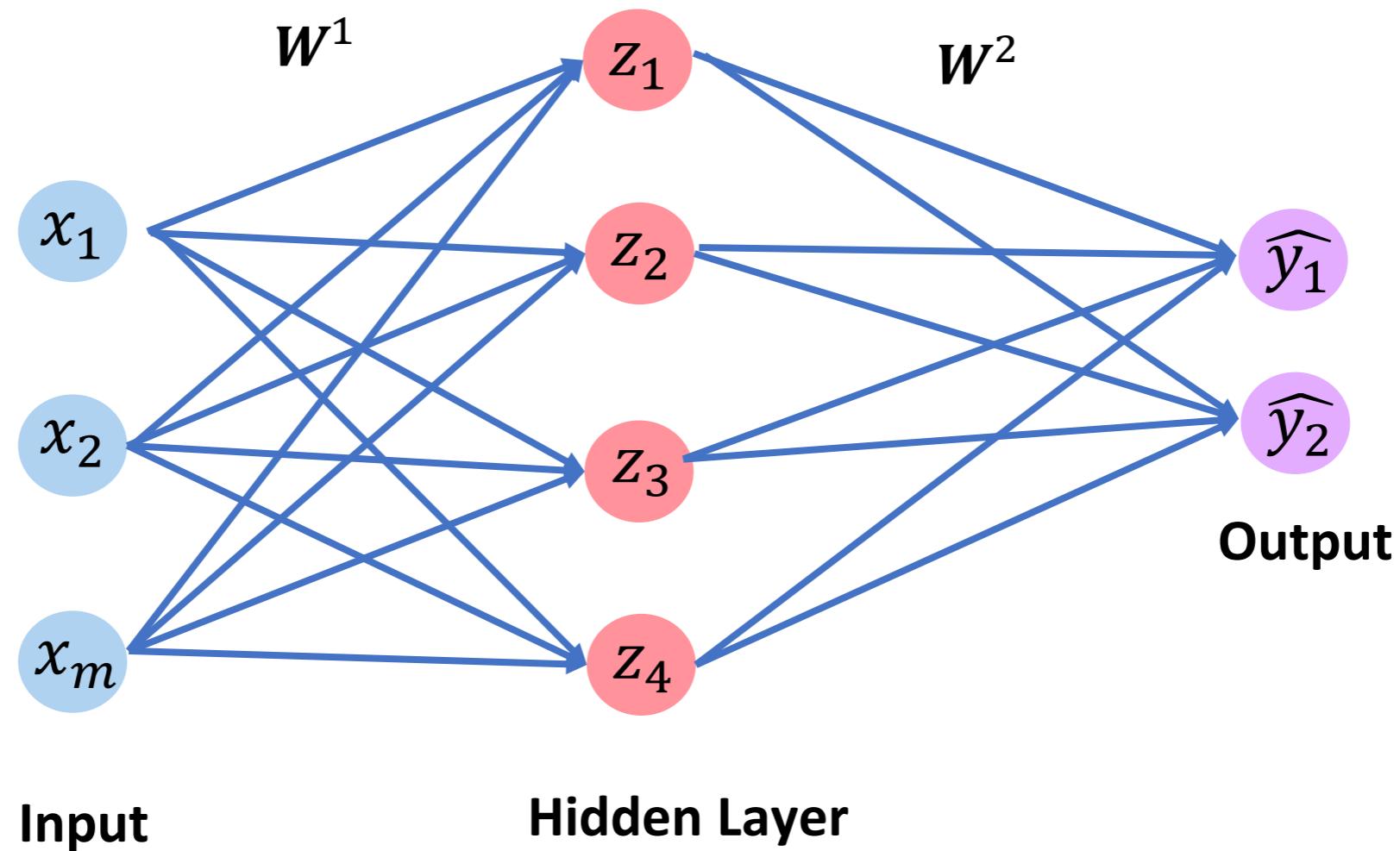
$$\hat{y} = g \left( \mathbf{w}_0 + \sum_{i=0}^m x_i w_i \right)$$

# Neural Network

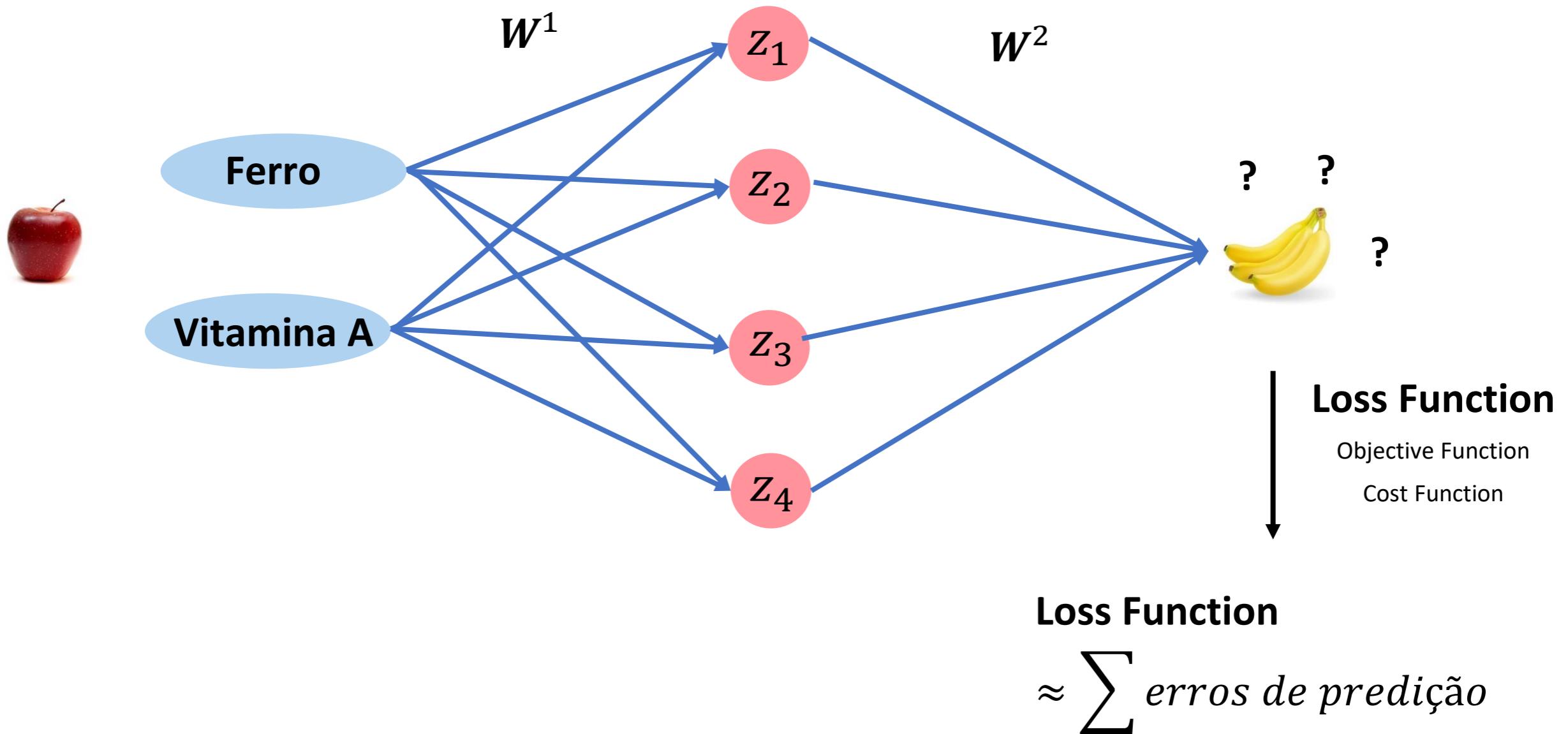


$$\hat{y} = g \left( \mathbf{w}_0 + \sum_{i=0}^m x_i w_i \right)$$

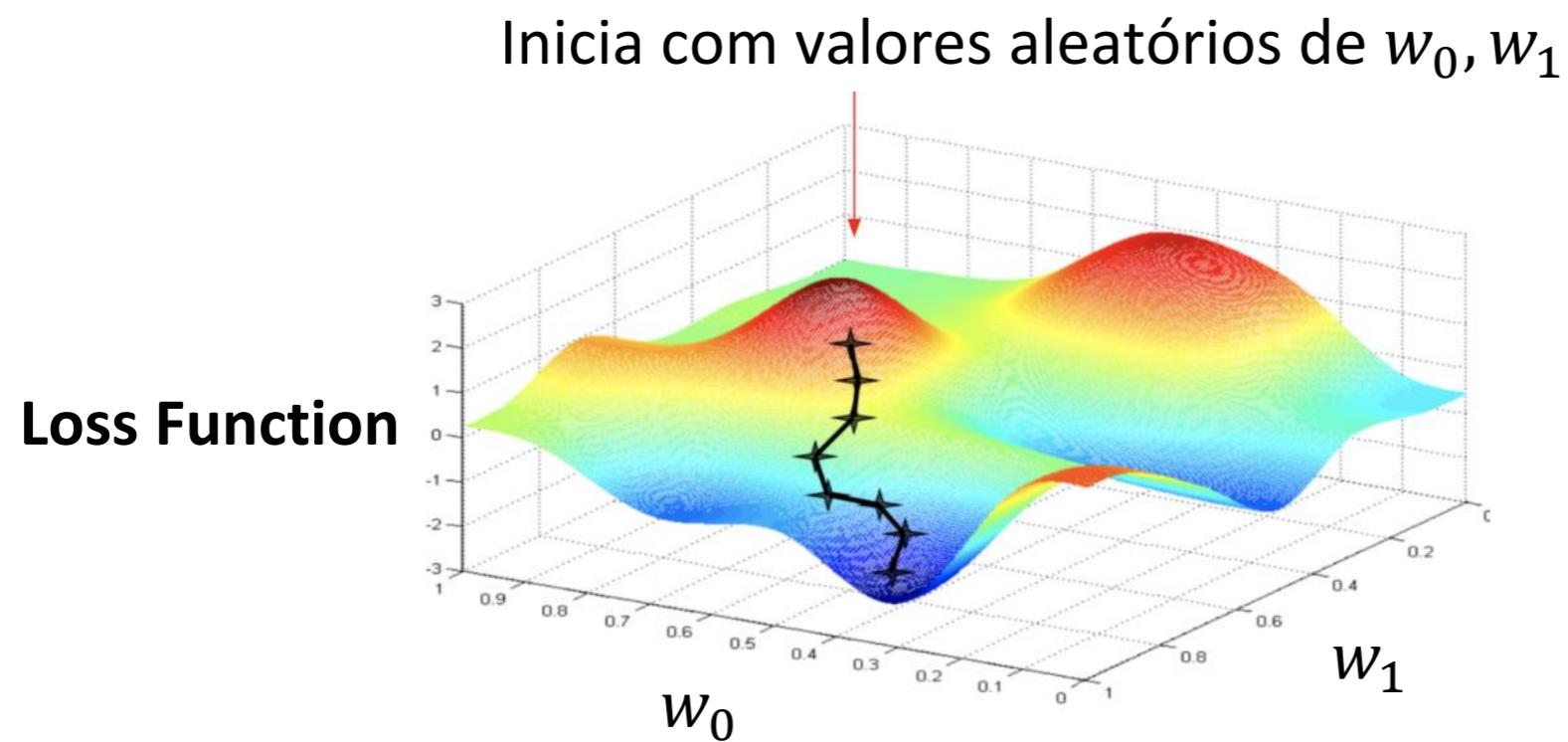
# Neural Network



# Neural Network

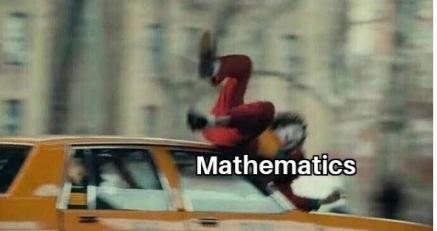
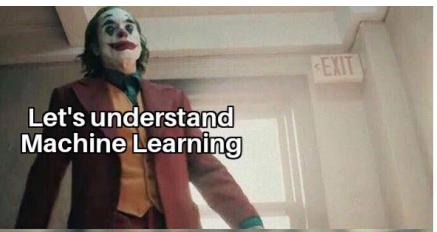


# Neural Network



<https://playground.tensorflow.org/>

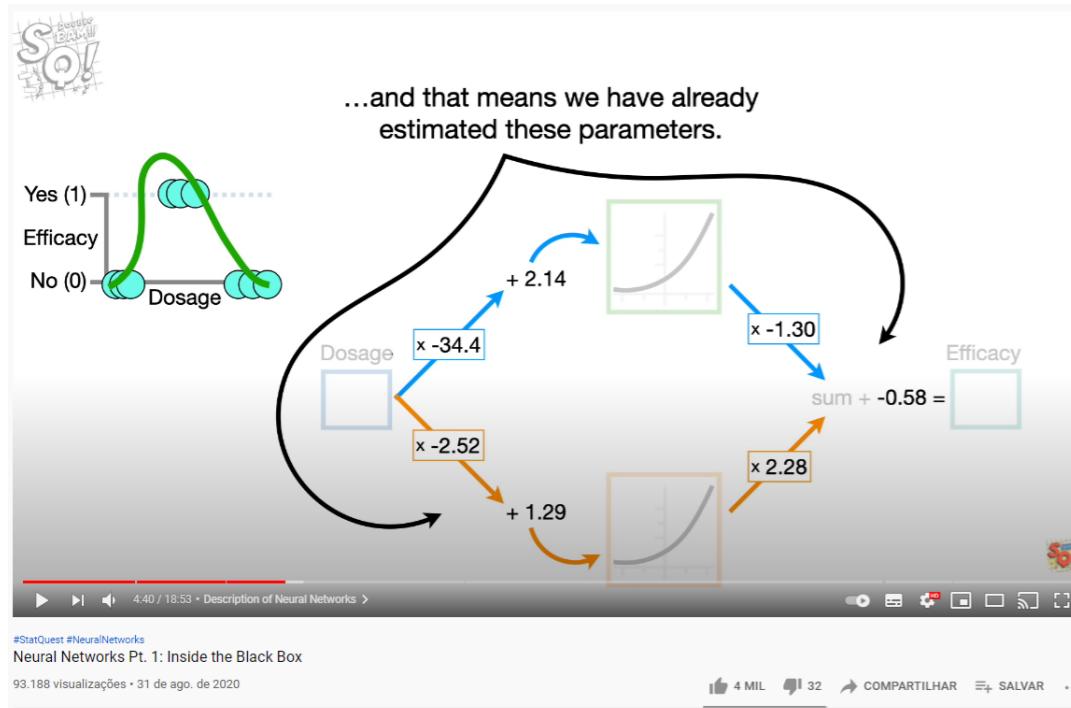




# Neural Networks

*Want to know more??*

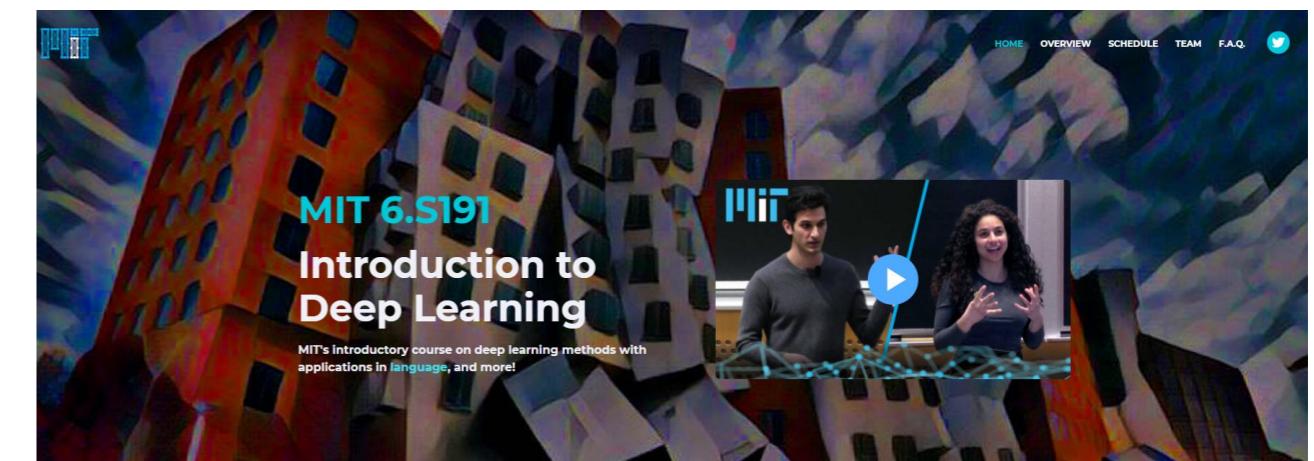
Easy mode



StatQuest with Josh Starmer

<https://youtu.be/CqOfi41LfDw>

Math and statistics required!!



MIT Introduction to Deep Learning

<http://introtodeeplearning.com/>

## Iago Costa

Geofísico e *Coordenador Executivo da Diretoria de Geologia e Recursos Minerais*

Serviço Geológico do Brasil – CPRM

e-mail: [iago.costa@sgb.gov.br](mailto:iago.costa@sgb.gov.br)

Telefone: 61-2108-8413

[www.cprm.gov.br](http://www.cprm.gov.br)

