

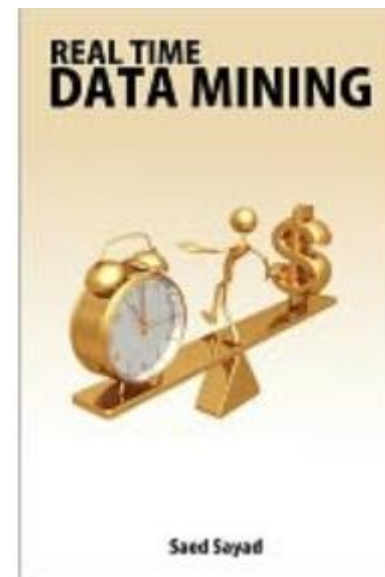
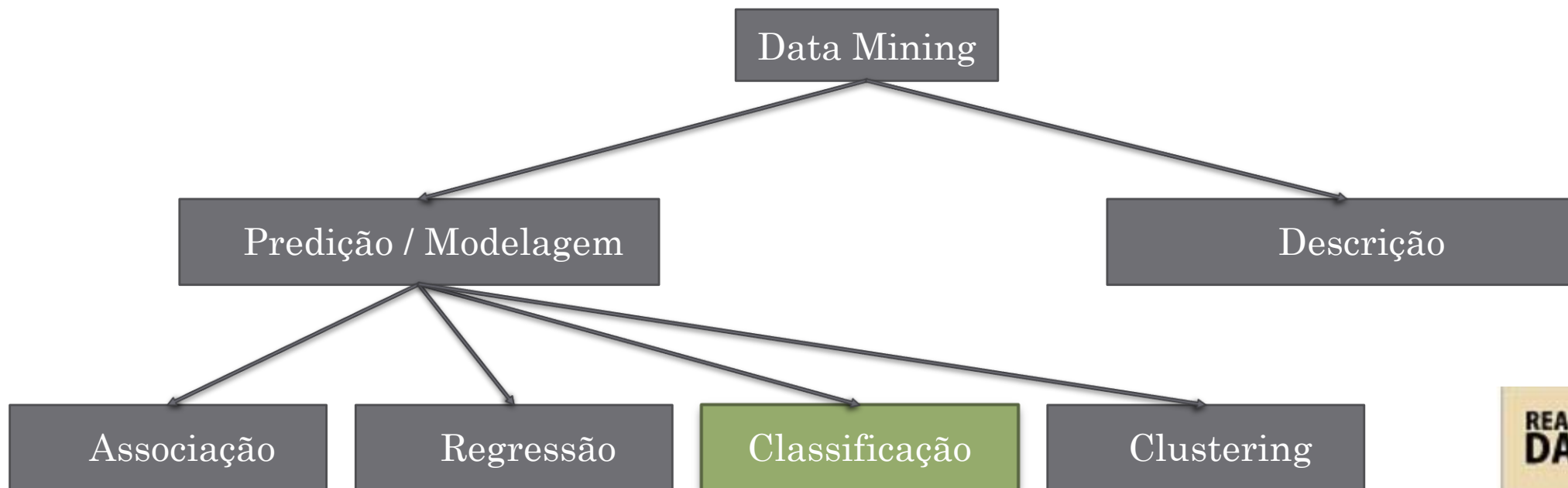
Data Mining

Regras de Classificação

Prof. Dr. Joaquim Assunção

DEPARTAMENTO DE COMPUTAÇÃO APLICADA
CENTRO DE TECNOLOGIA
UFSM
2019

Mapa para Mineração de Dados*



Notas legais

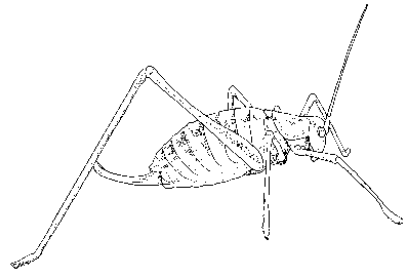
- Este material foi cedido pelo Dr. Eamonn Keogh (University of California - Riverside, US) para as aulas de mineração de dados na UFSM.
- Se você deseja usar este material para outros fins, entre em contato com o autor (eamonn@cs.ucr.edu).

Tradução e adaptação: Dr. Joaquim Assunção (joaquim@inf.ufsm.br).

Problema de Classificação

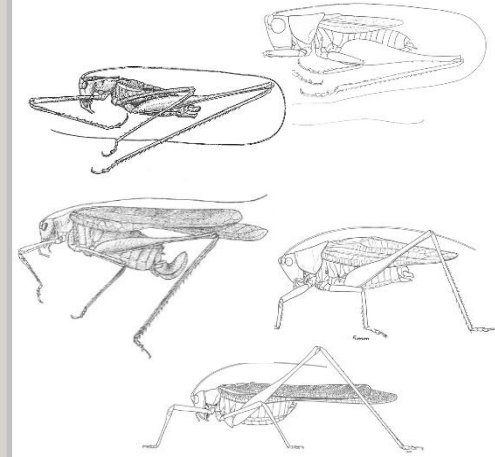
Uma definição informal

Dada uma coleção de dados anotados.
Neste caso, 5 instâncias de *Katydid*
(gafanhoto A) e 5 *Grasshopper* (gafanhoto B).
Decida que tipo de inseto o exemplo não
rotulado é.

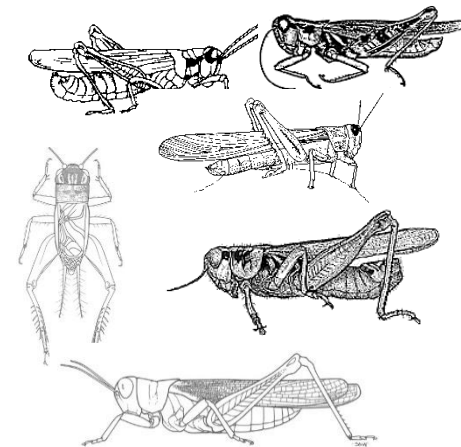


Katydid or *Grasshopper*?

Katydids



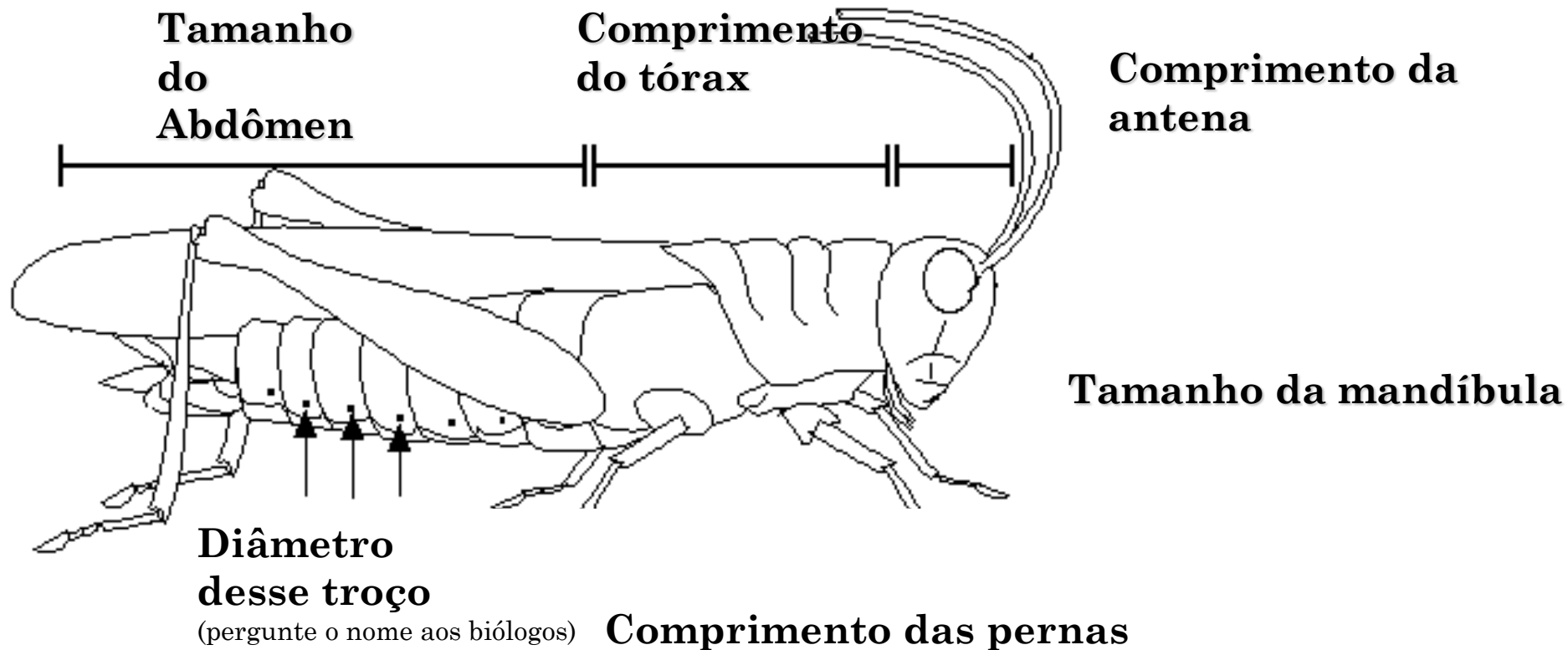
Grasshoppers



Para cada domínio de interesse, podemos medir as características

Cor {Verde, Cinza, Marrom, Outra}

Tem azas?



Podemos usar características em um conjunto de dados.

O problema de classificação agora pode ser expressado como:

- Dado um conjunto de treino (**Minha_Colecao**), Descubra o rótulo da **classe** de uma Instância nunca vista (inseto novo)

Minha_Colecao

Inseto ID	Abdômen	Antena	Classe do inseto
1	2.7	5.5	Grasshopper
2	8.0	9.1	Katydid
3	0.9	4.7	Grasshopper
4	1.1	3.1	Grasshopper
5	5.4	8.5	Katydid
6	2.9	1.9	Grasshopper
7	6.1	6.6	Katydid
8	0.5	1.0	Grasshopper
9	8.3	6.6	Katydid
10	8.1	4.7	Katydid

Inseto novo =

11

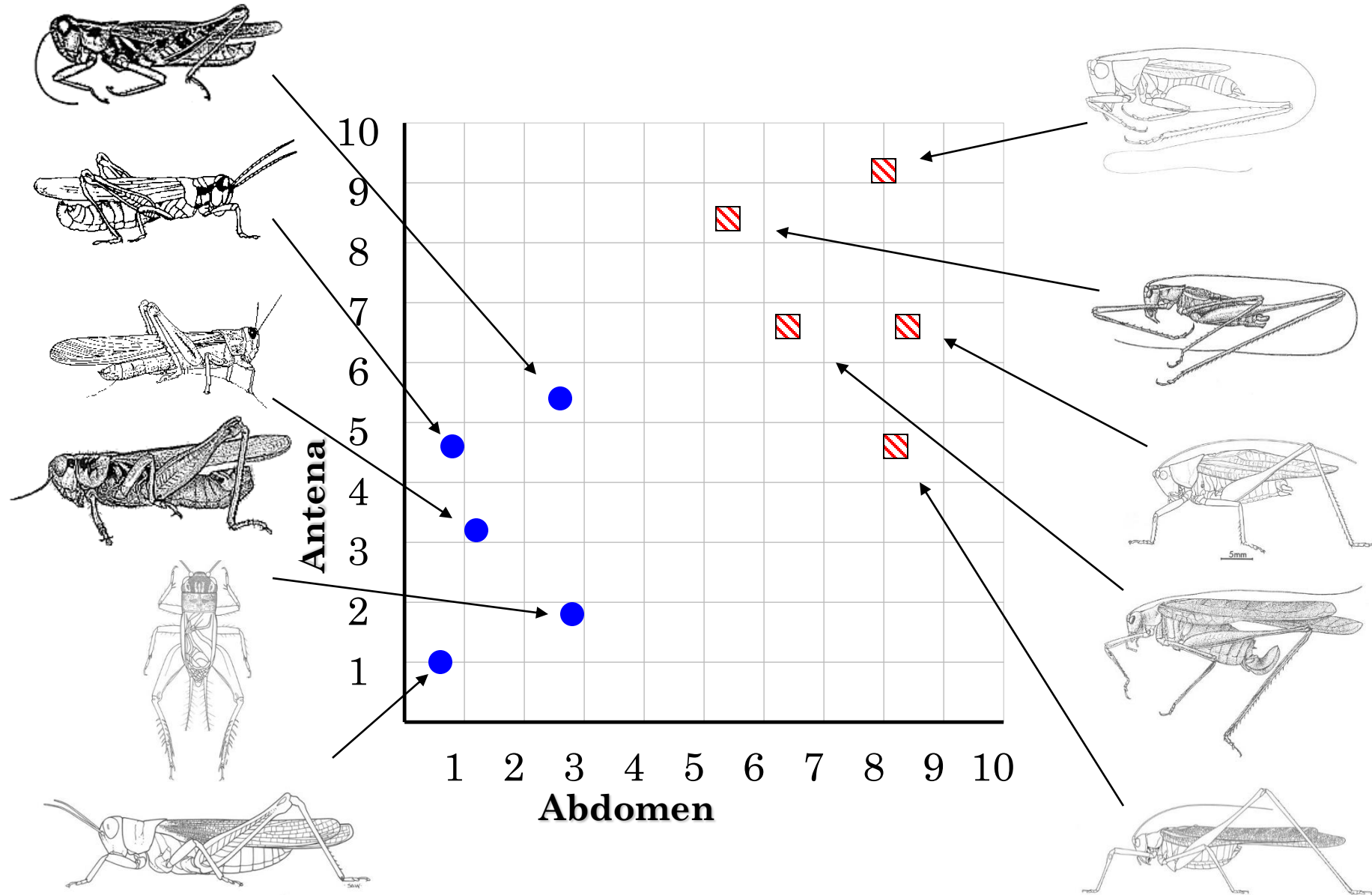
5.1

7.0

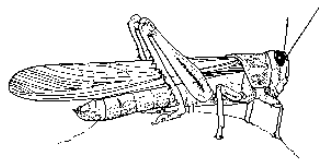
??????

Grasshoppers

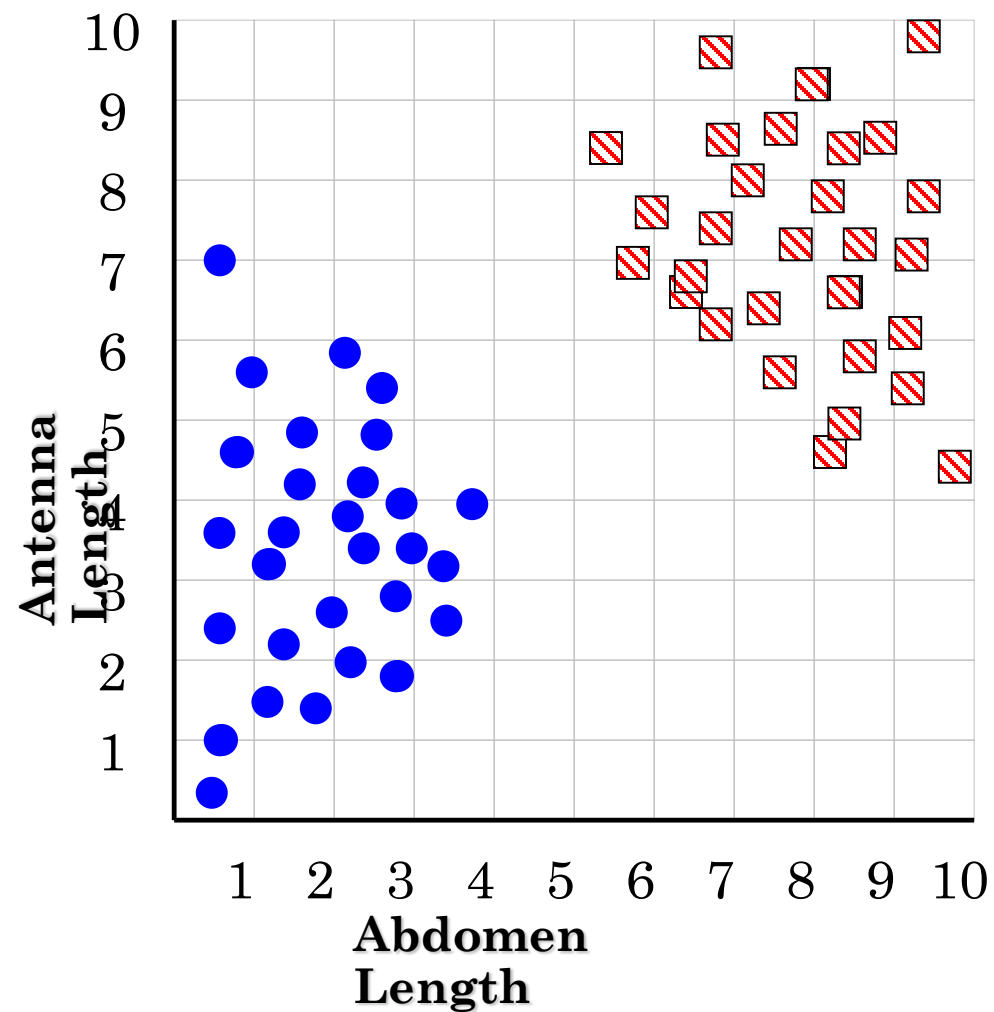
Katydids



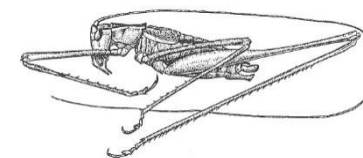
Grasshoppers



Vamos usar esse conjunto,
ligeiramente, maior como exemplo...



Katydid



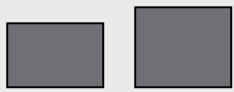
Cada um destes
objetos de dados
são chamados
de...

- exemplares
- exemplos de
treino
- instâncias
- tuplas

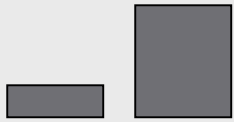
Voltaremos em breve.
Agora, vamos Jogar...

Jogo #1

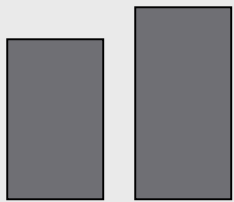
Exemplos da classe A



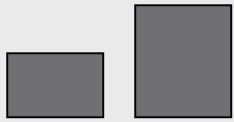
3 4



1.5 5

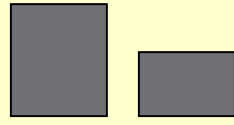


6 8

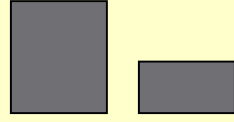


2.5 5

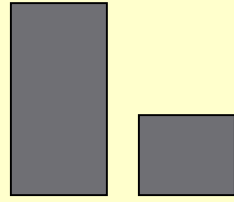
Exemplos da classe B



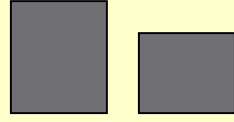
5 2.5



5 2



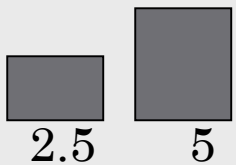
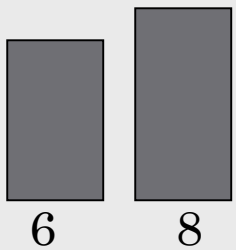
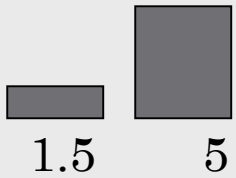
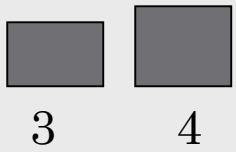
8 3



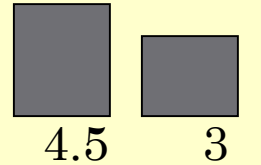
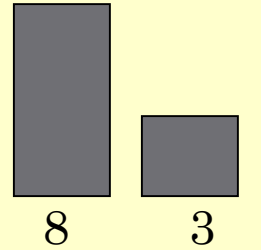
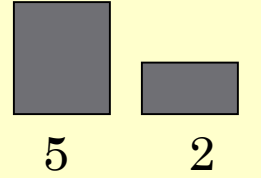
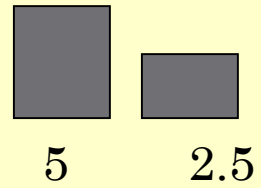
4.5 3

Jogo #1

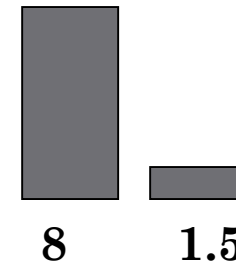
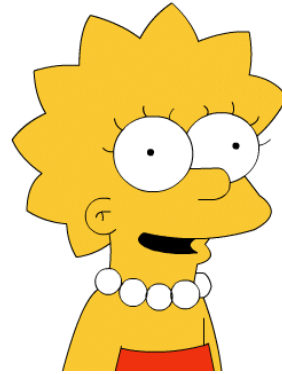
Exemplos da classe A



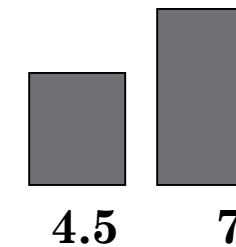
Exemplos da classe B



Que classe de objeto é essa?

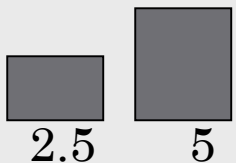
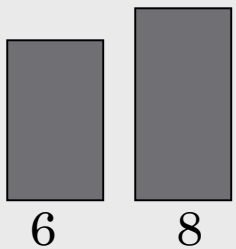
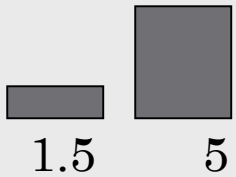
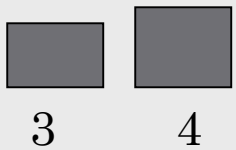


E essa, **A** or **B**?

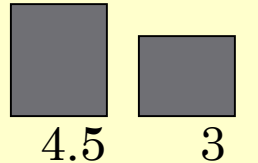
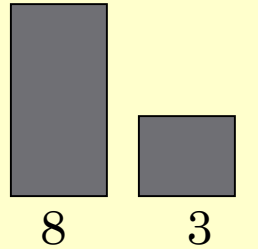
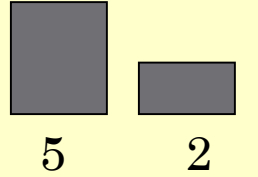
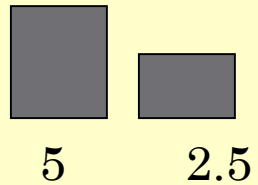


Jogo #1

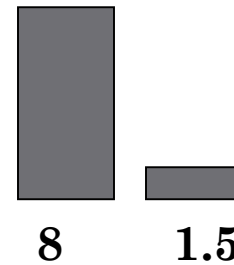
Exemplos da classe A



Exemplos da classe B



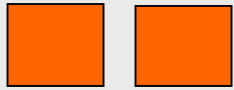
É **B**!



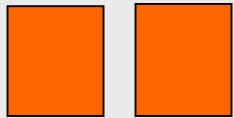
Aqui está a regra,
se o primeiro valor
for menor que o
segundo é **A**, caso
contrário é **B**.

Jogo #2

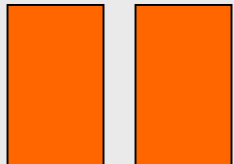
Exemplos da classe A



4 4



5 5

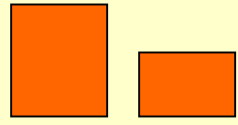


6 6

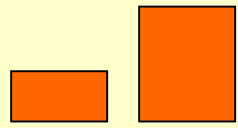


3 3

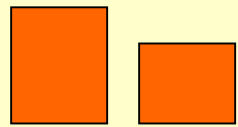
Exemplos da classe B



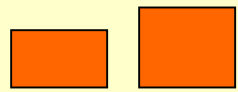
5 2.5



2 5

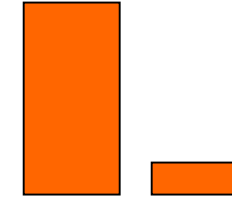


5 3



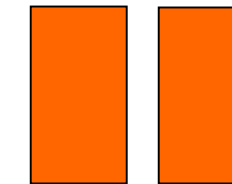
2.5 3

Humm! Éssa é difícil!



8 1.5

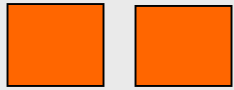
Até eu sei essa



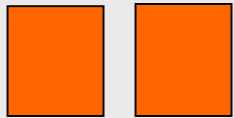
7 7

Jogo #2

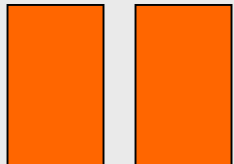
Exemplos da classe A



4 4



5 5

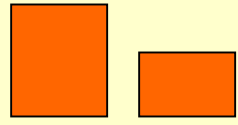


6 6

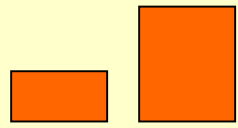


3 3

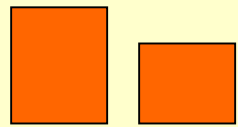
Exemplos da classe B



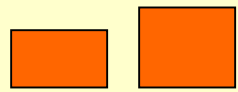
5 2.5



2 5



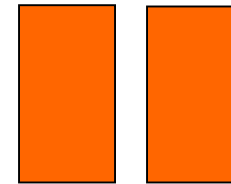
5 3



2.5 3

Aqui está a regra, se as barras forem de tamanhos iguais é um **A**. Caso contrário é **B**.

É um **A**! Sou um Gênio.



7 7

Jogo #3

Exemplos da classe A

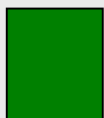


4

4



1



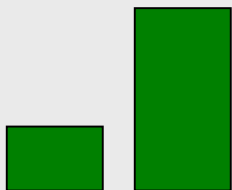
5



6



3

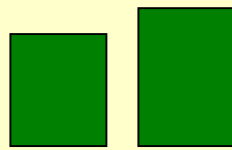


3



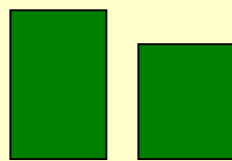
7

Exemplos da classe B



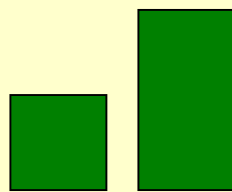
5

6



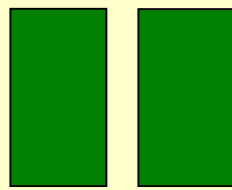
7

5



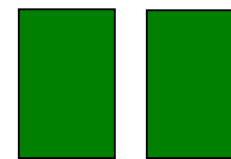
4

8



7

7



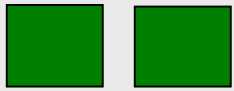
6

6

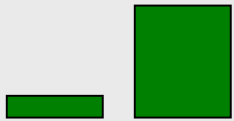
Putz, essa é realmente difícil...
A ou **B**?

Jogo #3

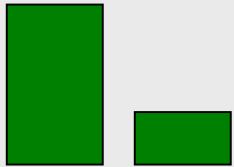
Exemplos da classe A



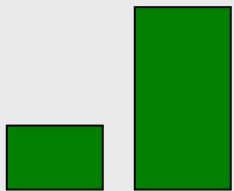
4 4



1 5

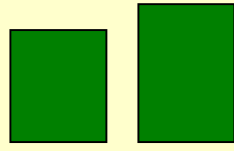


6 3

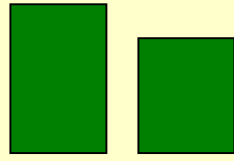


3 7

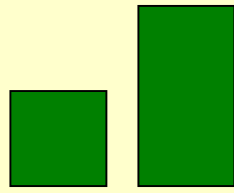
Exemplos da classe B



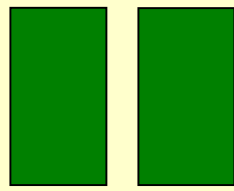
5 6



7 5

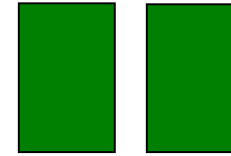


4 8



7 7

É um **B**!



6 6

A regra é a seguinte: Se o quadrado da soma das barras for menor ou igual a 100, é um **A**. Caso contrário é um **B**.

As regras fazem sentido, mas
não estou certo quanto a
exclusividade das mesmas...
vontade de assistir os Simpsons

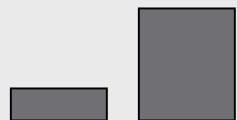


Jogo #1

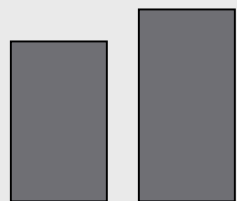
Examplos da classe A



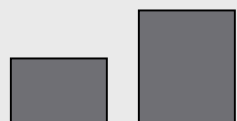
3 4



1.5 5



6 8

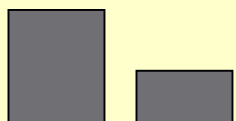


2.5 5

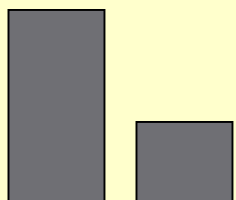
Examplos da classe B



5 2.5



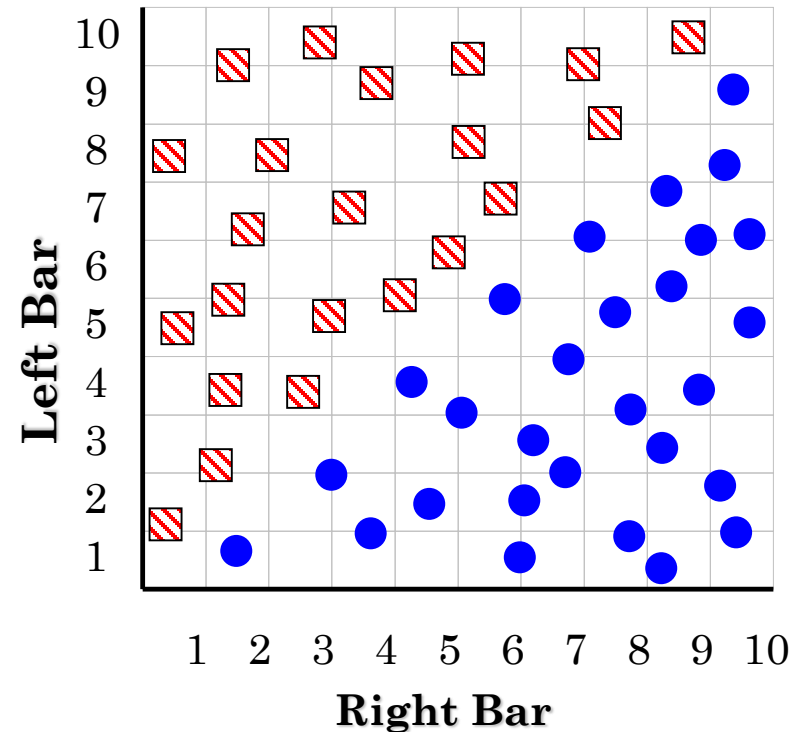
5 2



8 3



4.5 3



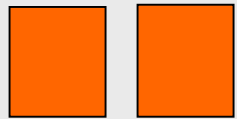
Relembrando a regra.
Se a barra esquerda for menor que a direita é **A**,
caso contrário é **B**.

Jogo #2

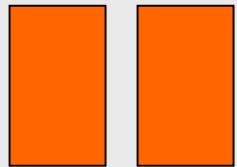
Exemplos da classe A



4 4



5 5

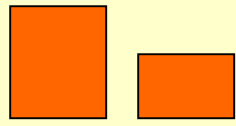


6 6

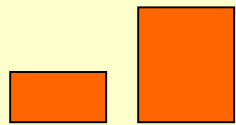


3 3

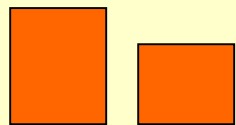
Exemplos da classe B



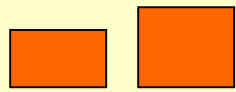
5 2.5



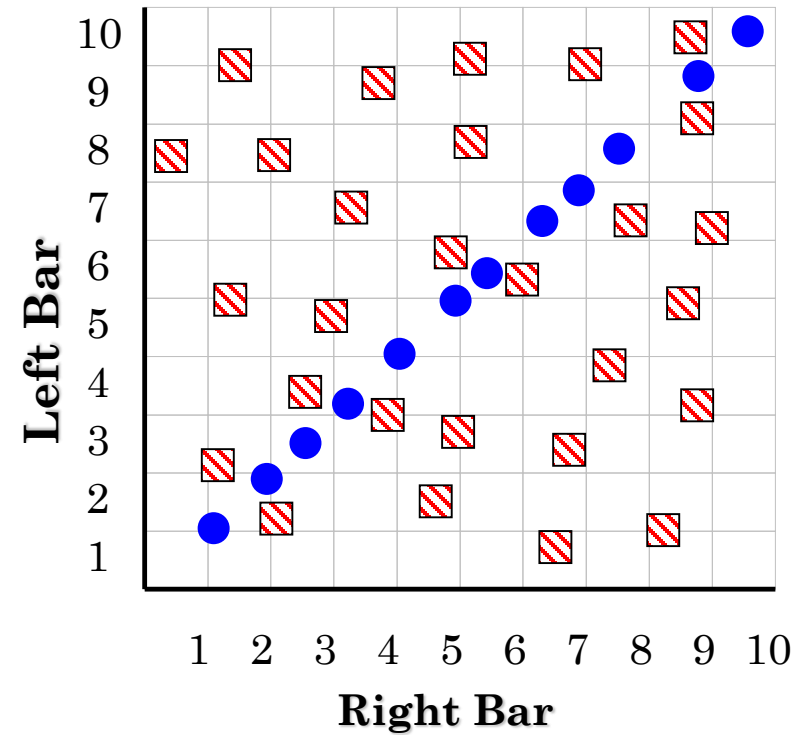
2 5



5 3



2.5 3

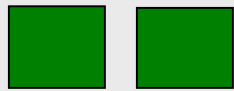


Deixa eu ver... aqui, achei! .. A regra é, se as duas barras tiverem o mesmo tamanho é **A**. as outras são **B**.



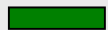
Jogo #3

Exemplos da
classe A



4

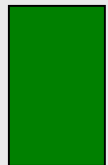
4



1



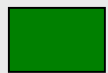
5



6



3

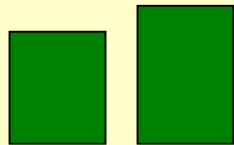


3



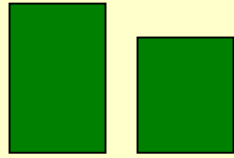
7

Exemplos da
classe B



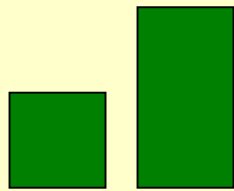
5

6



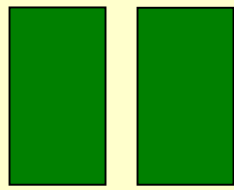
7

5



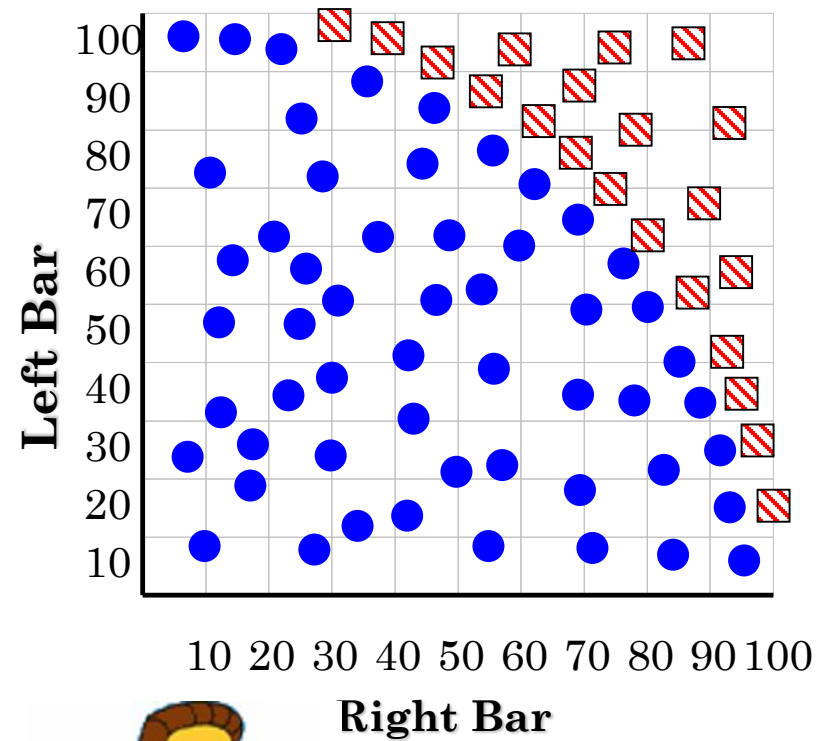
4

8



7

7



A regra é:
Se o quadrado da soma das
barras é menor ou igual a 100 é
A. Caso contrário é **B**.

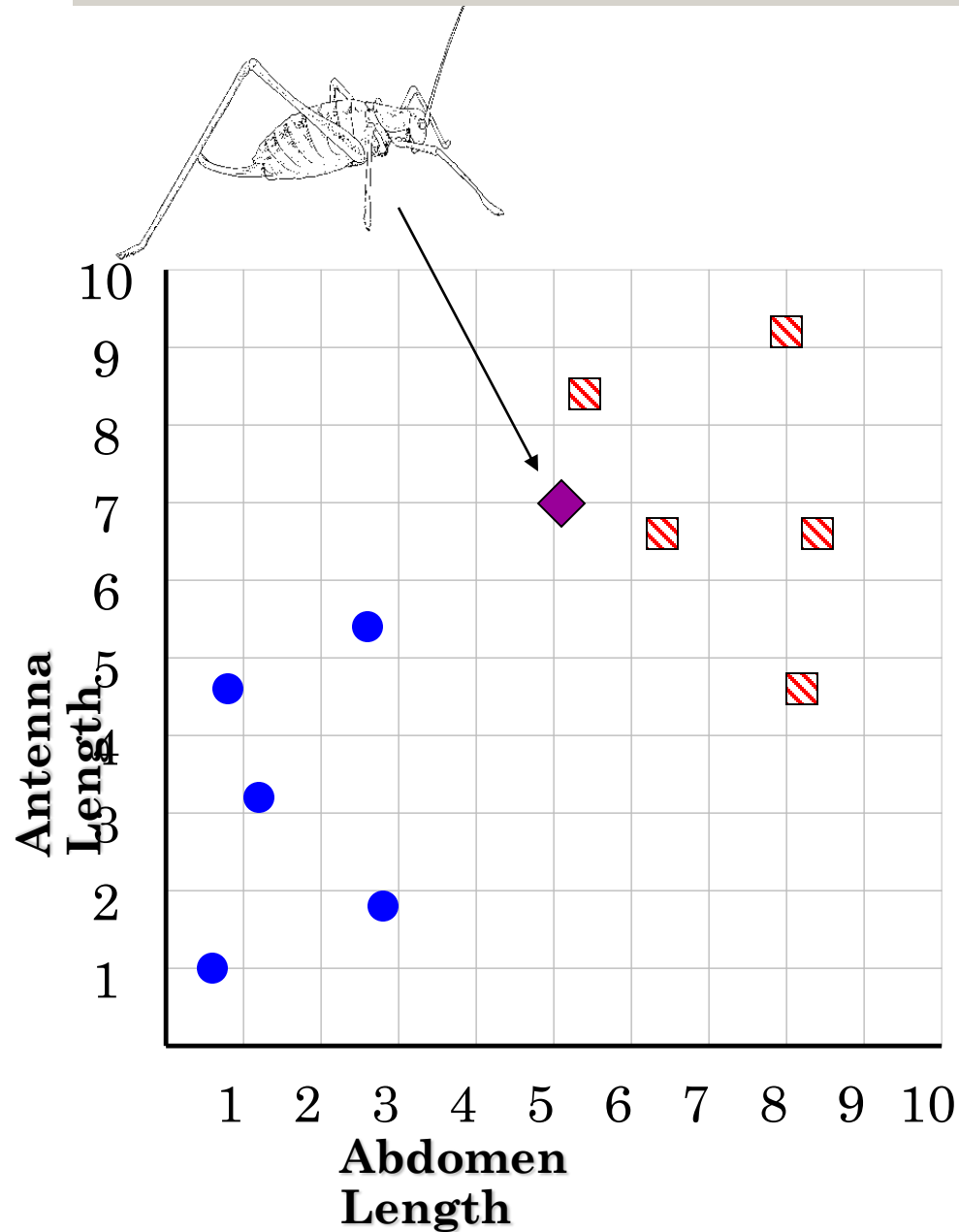
Inseto desconhecido →

11

5.1

7.0

???????



- Podemos projetar o **inseto desconhecido** no mesmo espaço dos demais dados do conjunto.
- Agora que abstraímos os detalhes, será mais fácil falar dos dados.

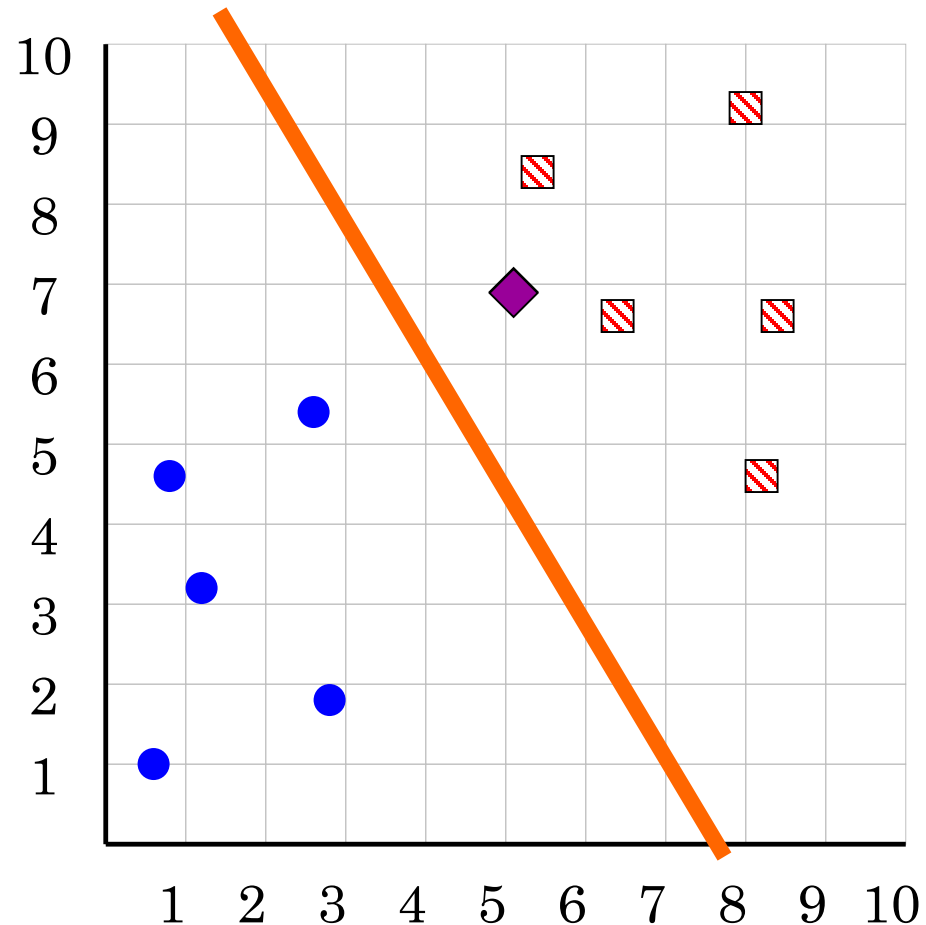
▣ **Katydid**

● **Grasshoppers**

Classificador linear simples



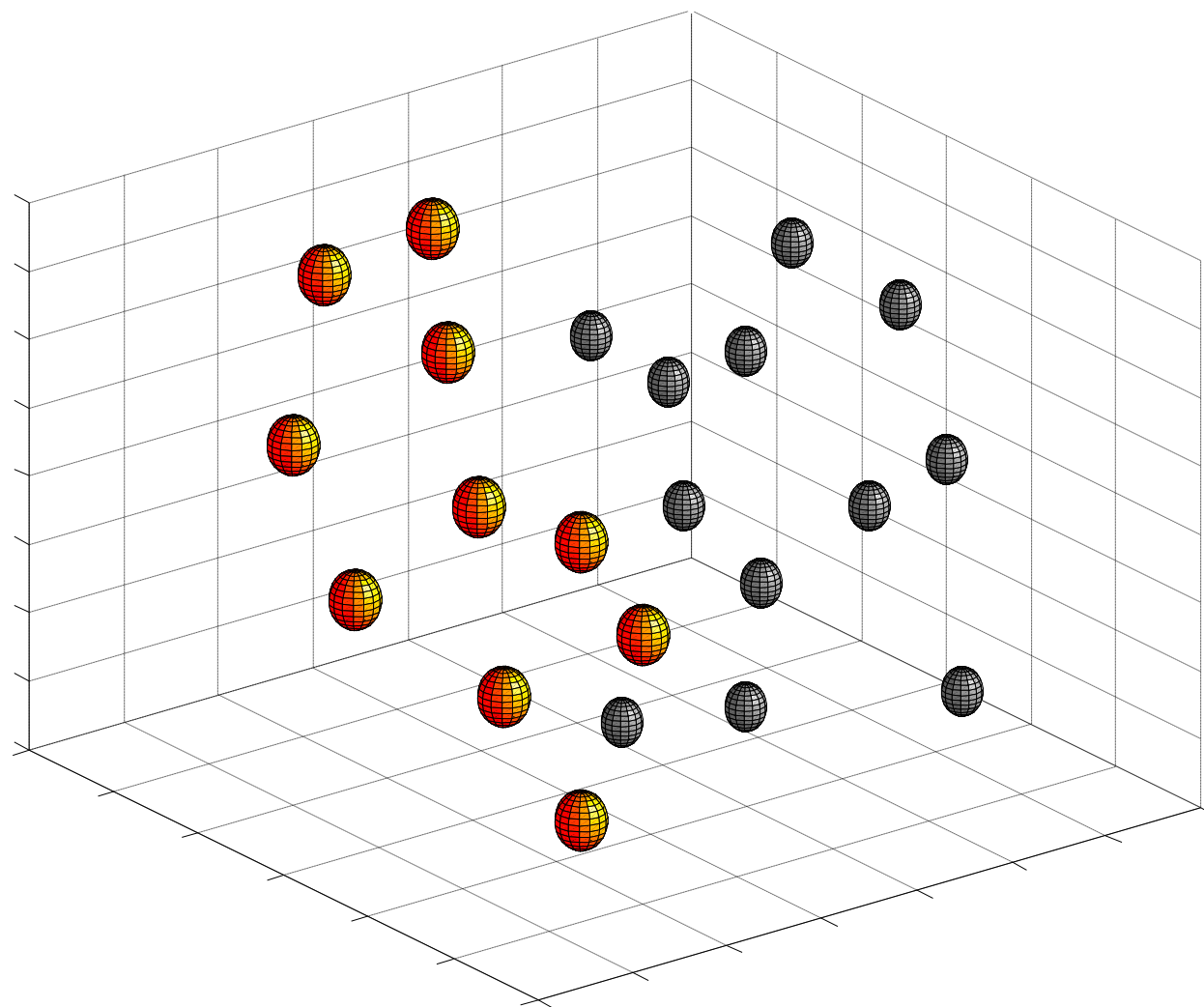
R.A. Fisher
1890-1962



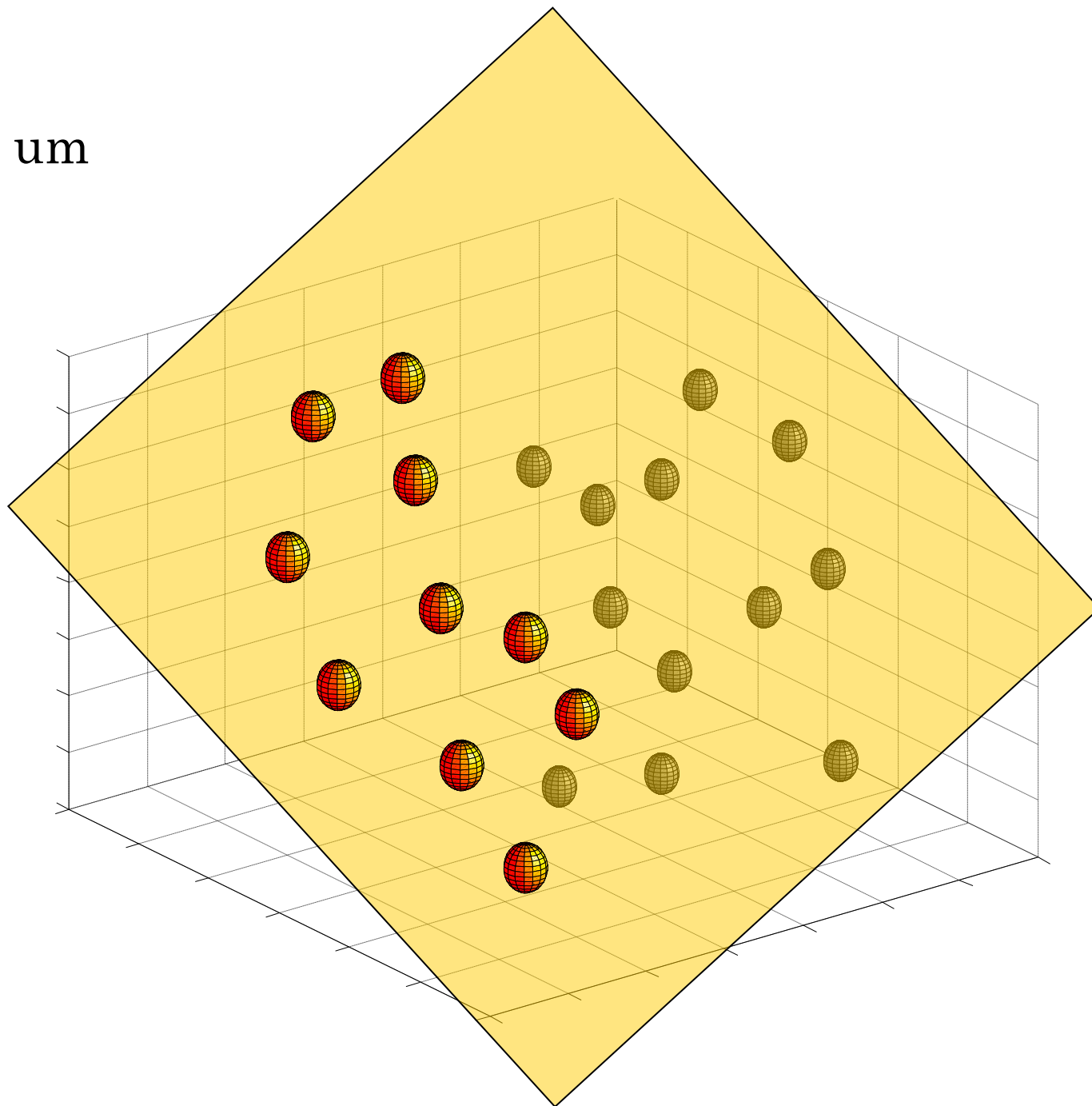
Se o atributo desconhecido
está acima da linha
então
a classe é **Katydid**
senão
a classe é **Grasshopper**

▣ **Katydid**
● **Grasshoppers**

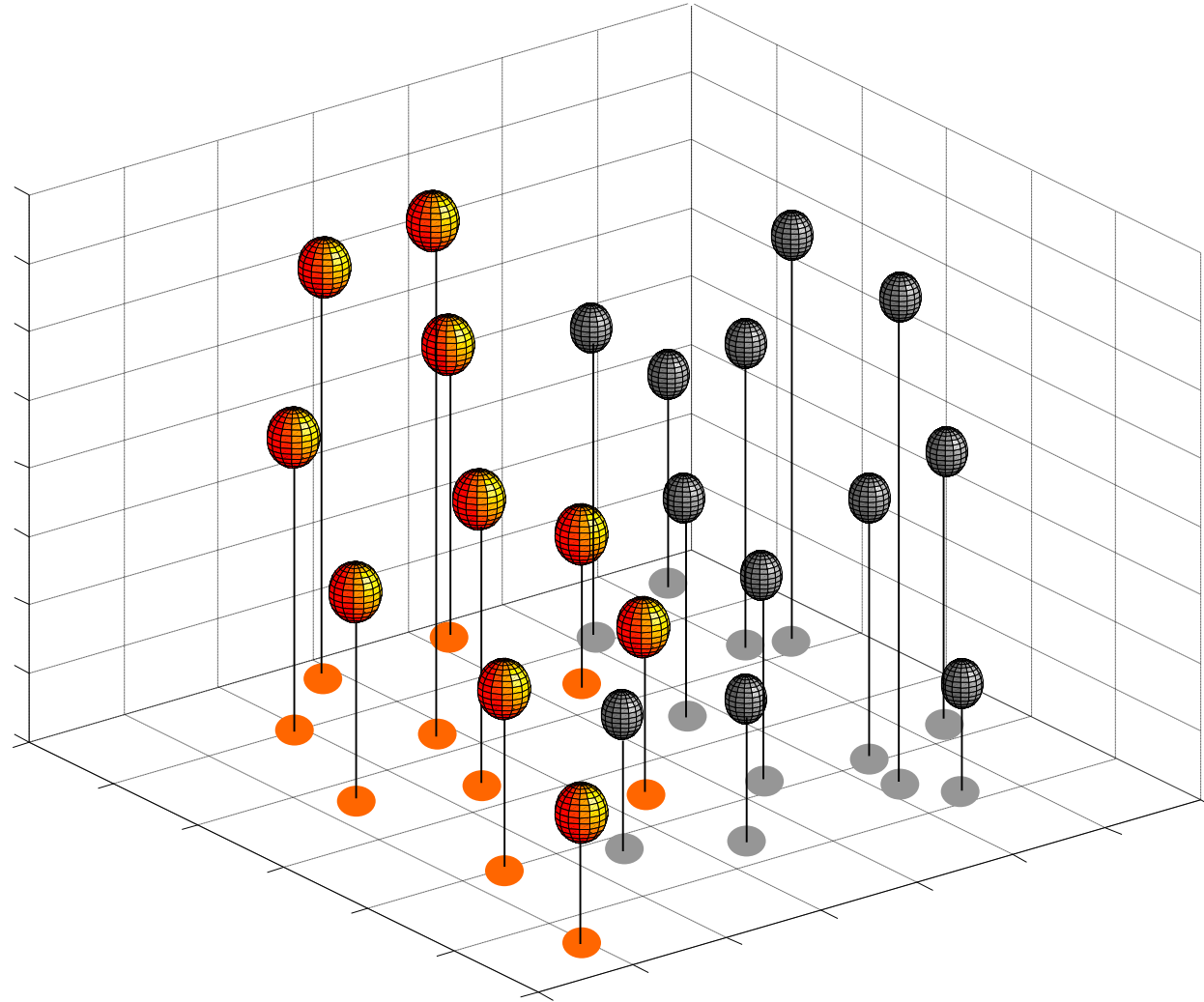
Esse mesmo tipo de
classificador pode ser usado
em
mais espaços
dimensionais...



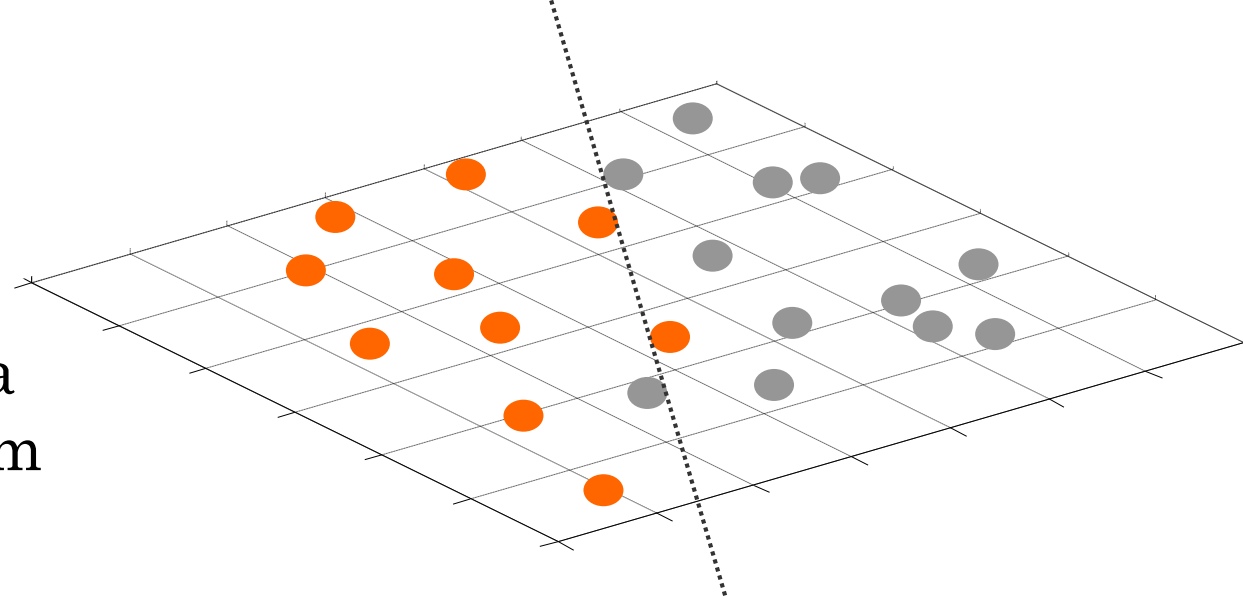
... Podemos
visualizar como um
n-dimensional
hiperplano.



Podemos pensar
o que aconteceria
neste mesmo
exemplo se não
tivéssemos a
Terceira
dimensão...

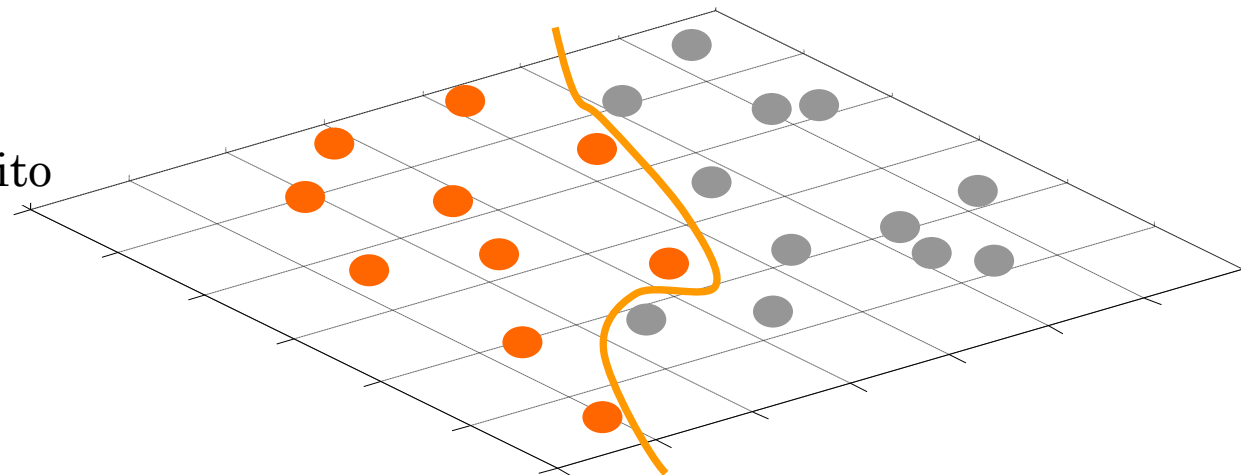


Já não podemos ter uma
acurácia perfeita com um
classificador linear...



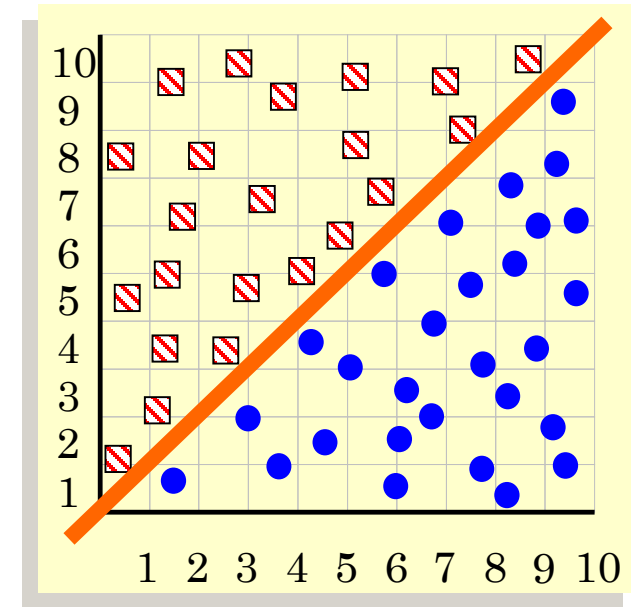
Poderíamos tentar solucionar
esse problema com um
classificador quadrático ou
cúbico...

No entanto, não é uma ideia muito
boa...

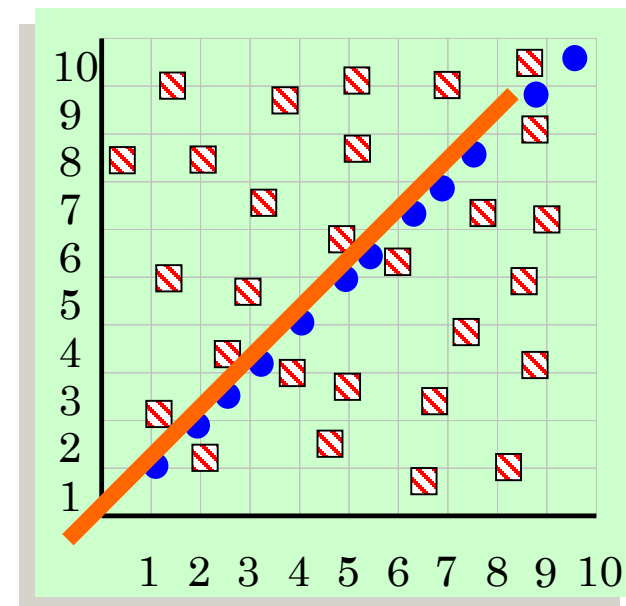
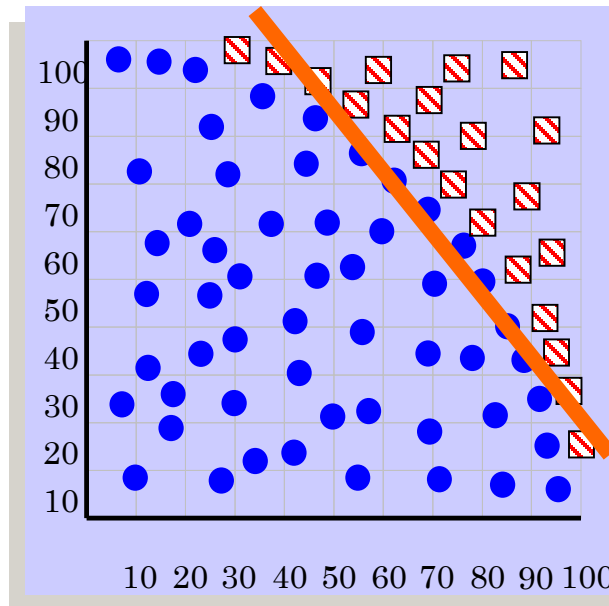


Quais dos #jogos podem ser resolvidos com um classificador linear simples?

- 1) Perfeito
- 2) Inútil
- 3) Muito bom



Problemas que podem ser resolvidos por um classificador linear são chamados de **Separáveis linearmente**.



Hands On!

- Use `read.csv` para ler o arquivo “insetos00.csv”. Na sequência defina um modelo linear simples para classificar os insetos com as seguintes características como Grasshoper e katydid (respectivamente):

Abdômen= 3.2 e Antena=4.2

Abdômen= 7.2 e Antena=4.1

Tech help

Use `abline(y, inclinacao)` para definir sua linha