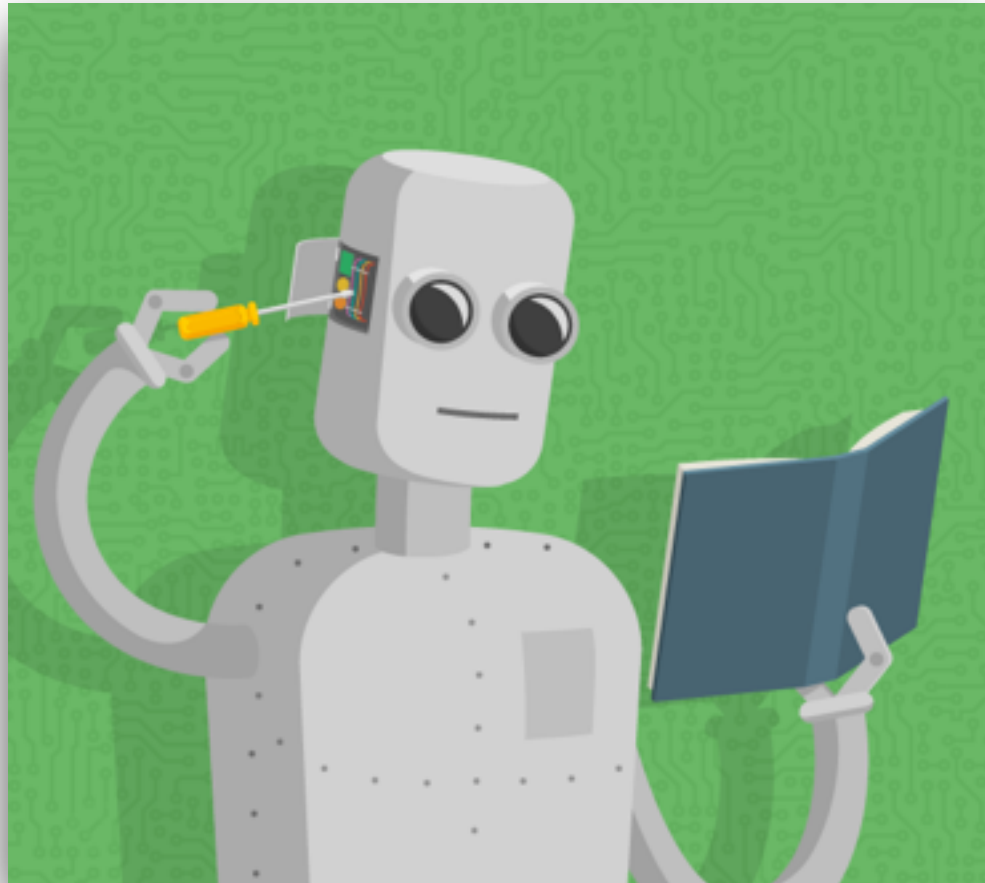




UNIFOR
ENSINANDO E APRENDENDO

Disciplina de Inteligência Artificial (Aula 3)



Classificação

Apresentação

Francisco Nauber Bernardo Gois



Analista aprendizado de máquina
no Serviço Federal de Processamento
de Dados

Doutorando em Informática Aplicada

Mestre em Informática Aplicada

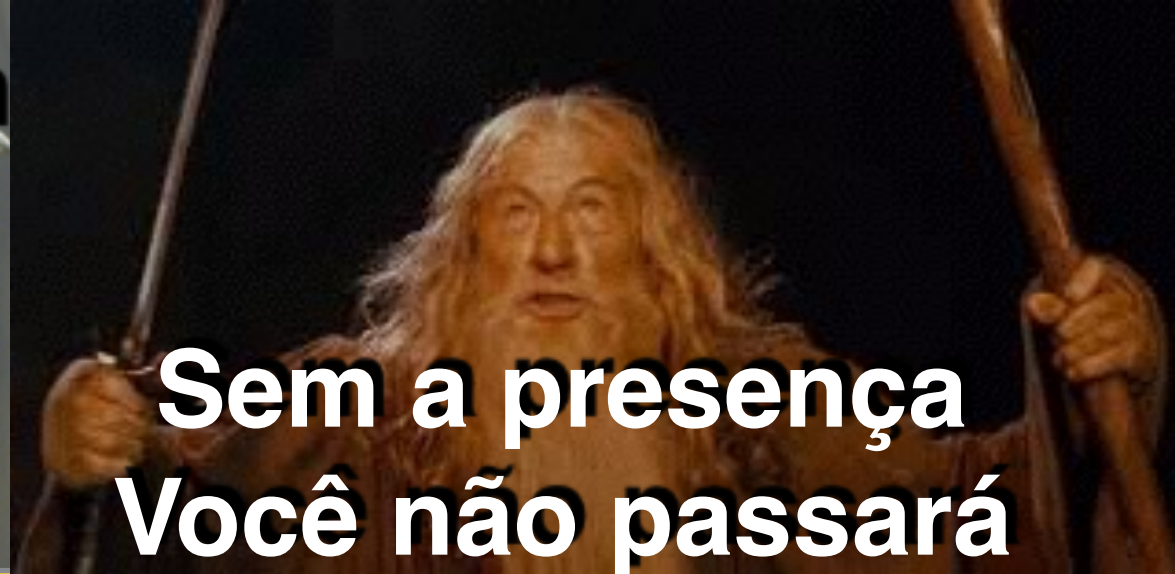
Especialista em desenvolvimento WEB

**Jovem Padawan
procure na aula**



**ao telefone
não falar**

**Sem a presença
Você não passará**



**Para melhor
desempenho na
aula**

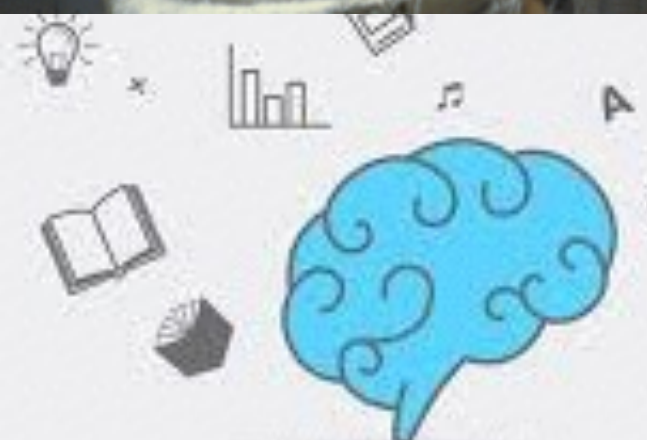
**Cuidado
com o
Horário**

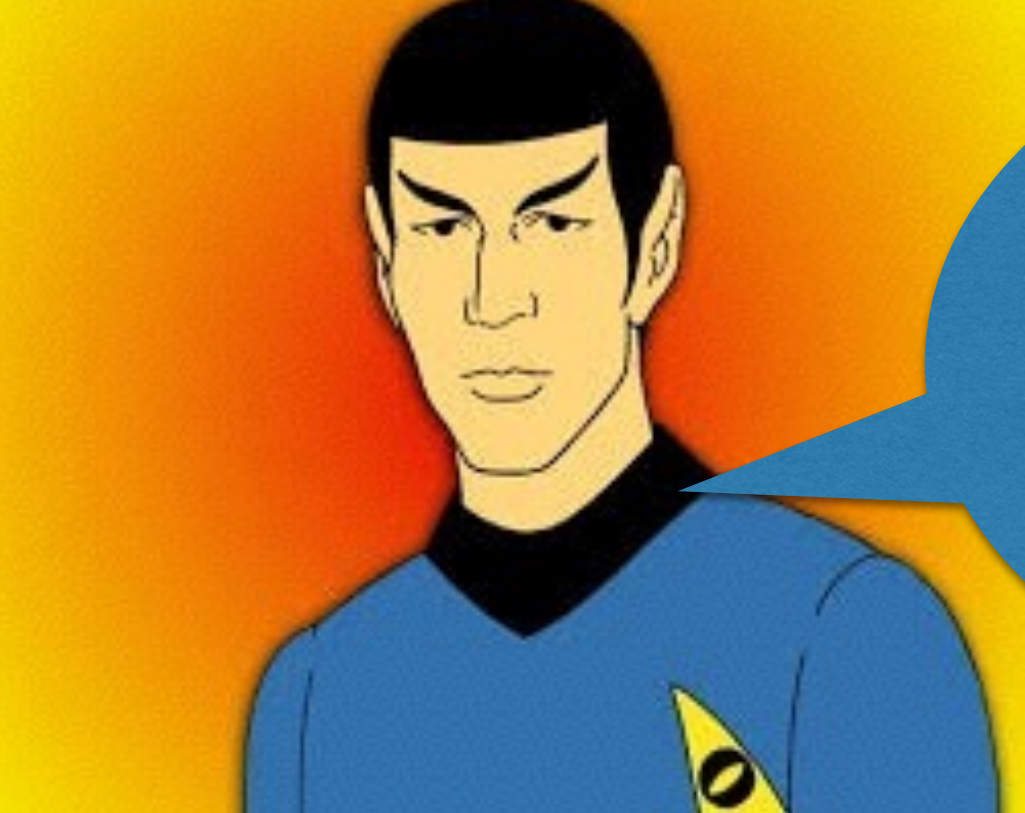
**Não
teremos
pontuação fora dos
trabalhos e provas
da disciplina**

**Buscar
aprendizado
ao invés de
pontos**



**Procure
não
conversar
durante a
aula**





OS trabalhos
deveram ser
entregues uma
semana antes
da prova

Um cadeira longa
e prospera

Não teremos
pontos após a prova
não adianta pedir



Cronograma da Disciplina



**Introdução
ao Python**



Classificação

Regressão



O que vimos na aula passada

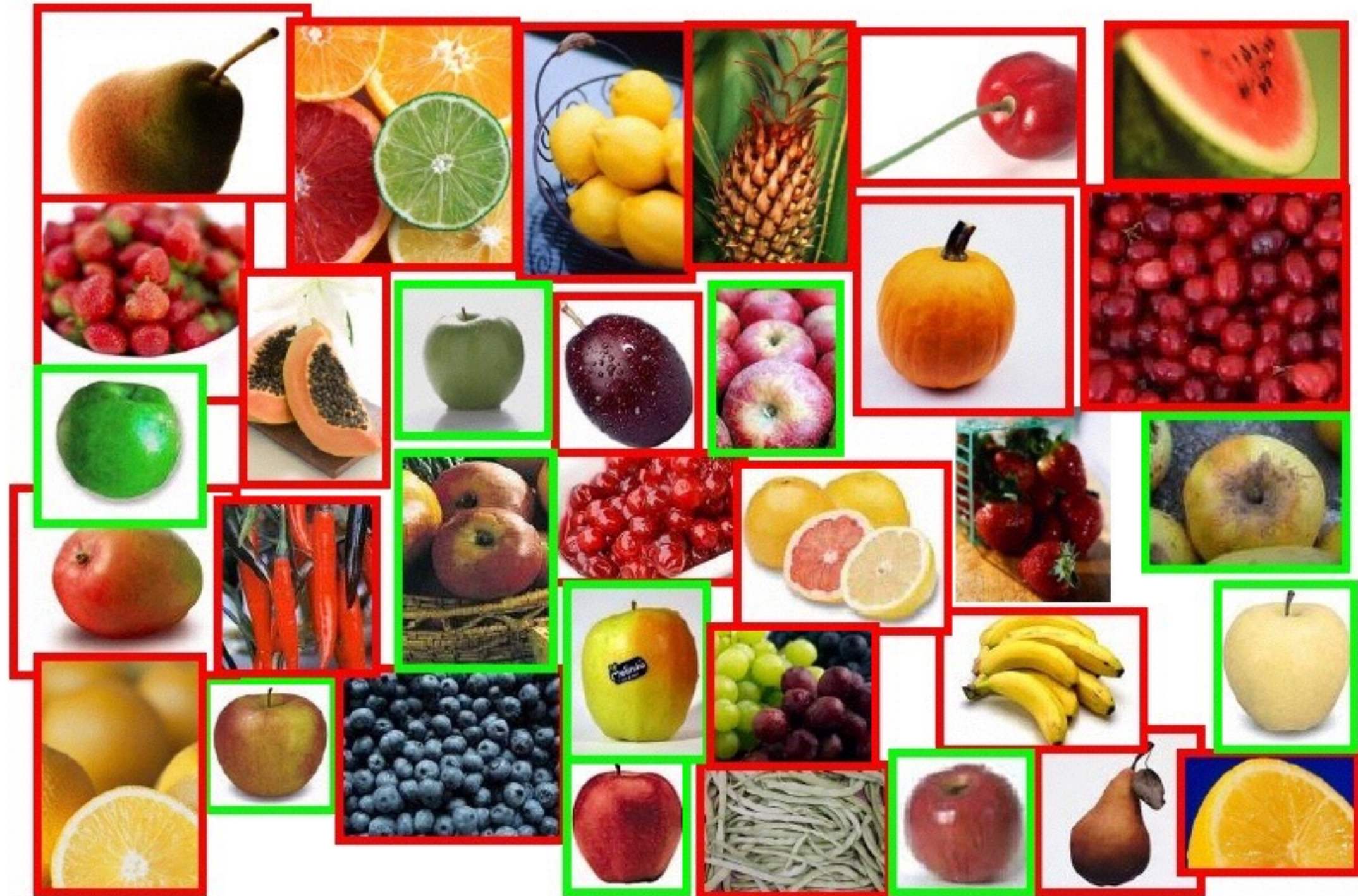
- **Introdução ao Python e ao Numpy**

Objetivo da Aula

- **Aprendizado de Máquina**

What is an apple?

3

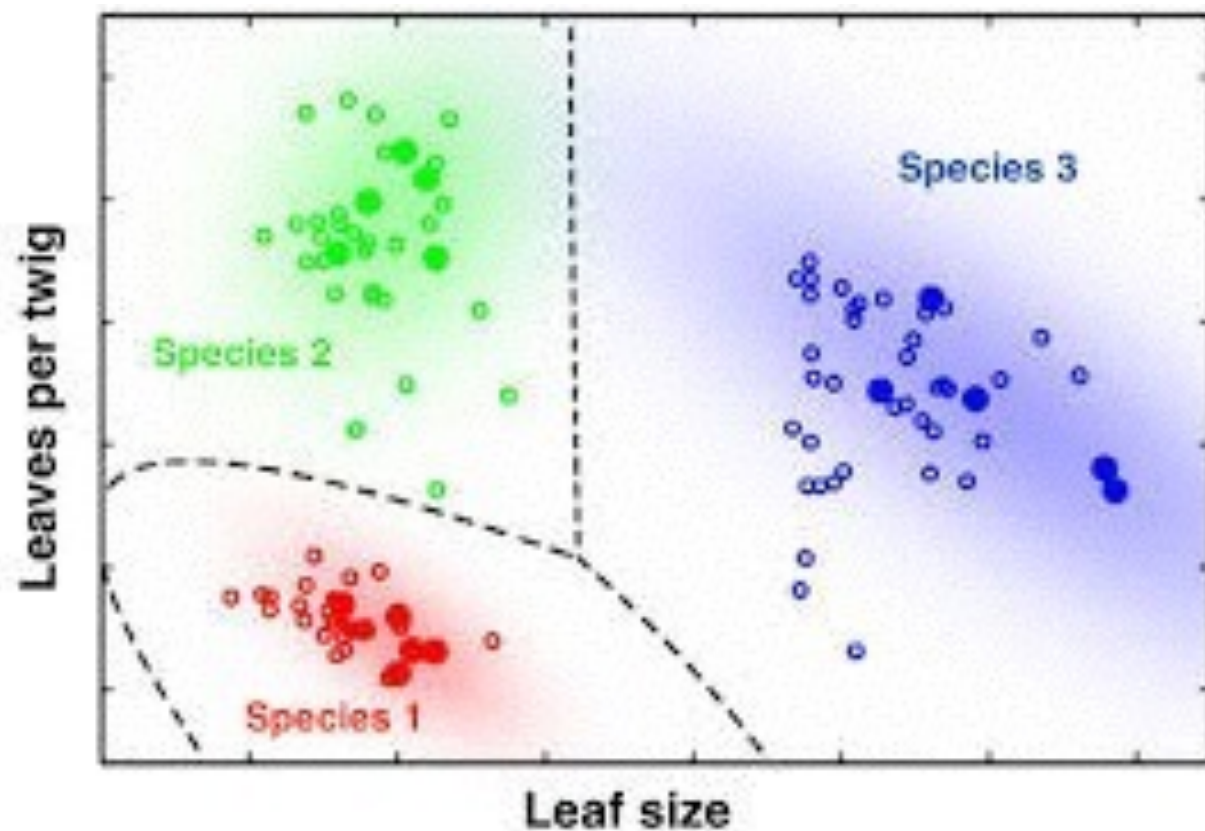


Are these apples?

4



Classificação



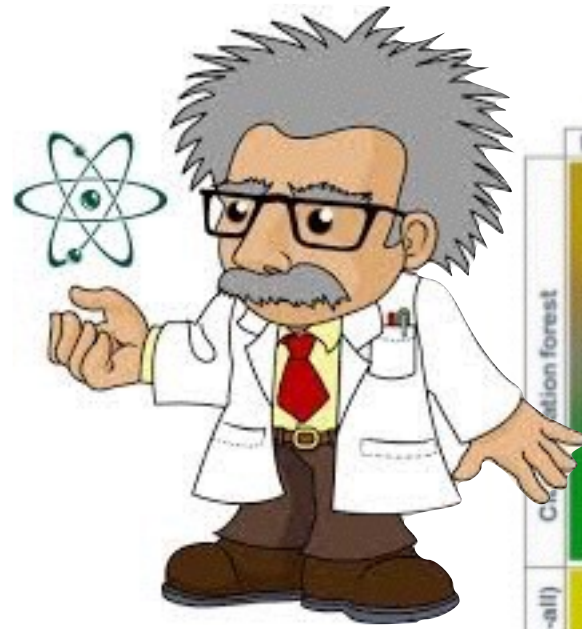
Classifica elementos partir de dados de treinamento

Classificação**Classificação**

- Analisa um conjunto de dados de treinamento (um conjunto de objetos cuja classificação já é conhecida) e constrói um modelo para cada classe baseado nas características dos dados. Uma árvore de decisão ou um conjunto de regras de classificação é gerado por tal processo de classificação, que pode ser usado para entender melhor cada classe ou classificar futuros dados. Na classificação o atributo a ser predito é discreto, por exemplo, se o cliente é bom ou mau pagador.

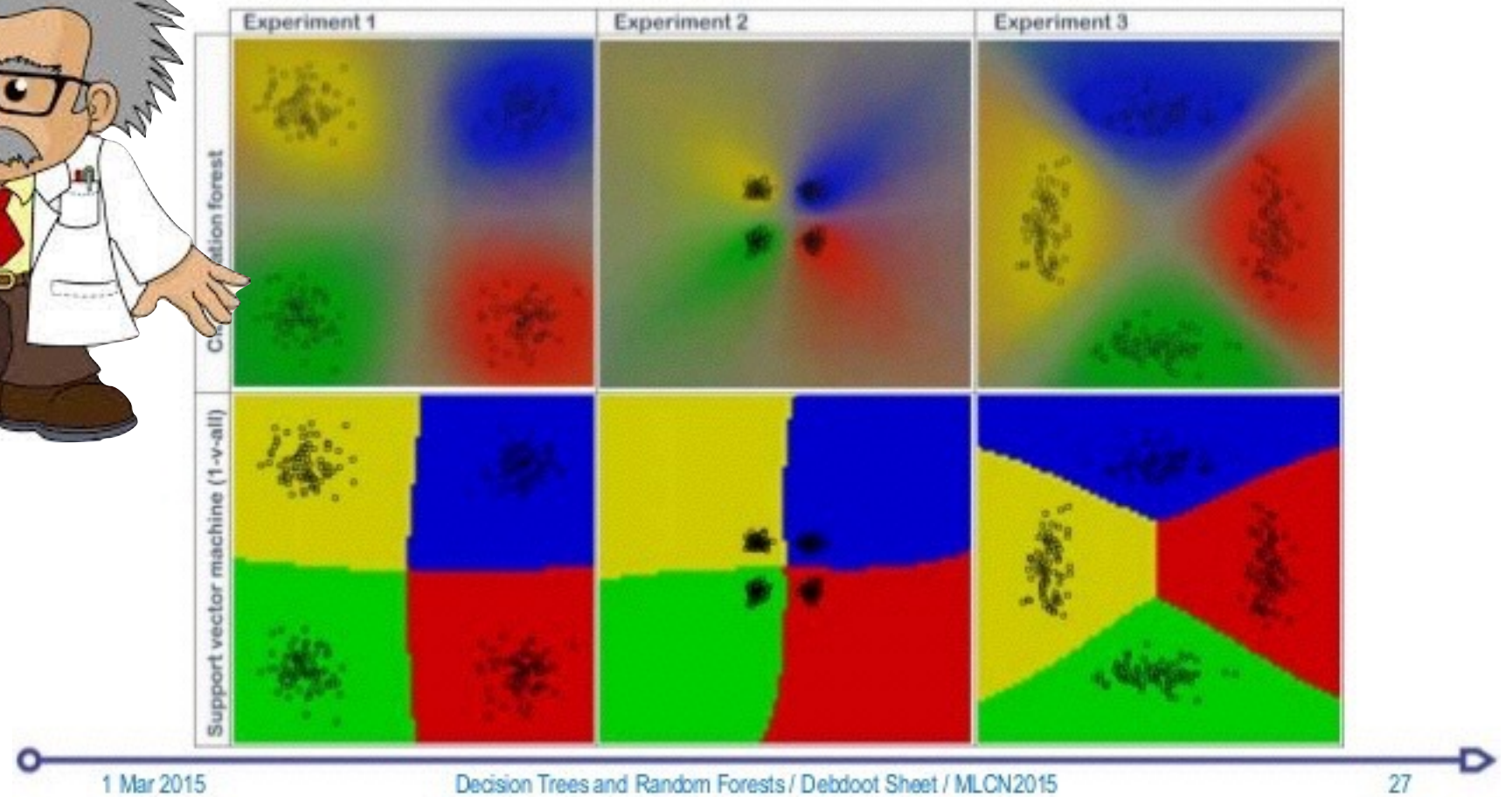
Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

Classificação



**Diferentes
algoritmos
cobrem
Diferentes
Superfícies**

Random Forest vs. SVM



Classificação

Descoberta de Conhecimento

- O problema é que estes registros da base dados, muitas vezes, representam apenas dados e não conhecimento. Visando transformar estes dados em conhecimento, surge o processo chamado de Descoberta de Conhecimento em Bancos de Dados, conhecido com KDD (Knowledge Discovery in Databases)
- A KDD é uma área de grande pesquisa na Inteligência Artificial.



Classificação**Aprendizado de Máquina**

- Um sistema de Aprendizado de Máquina (AM) é um programa de computador que toma decisões baseado em experiências acumuladas por meio de soluções bem-sucedidas de problemas anteriores.
- O AM é uma poderosa ferramenta para aquisição automática de conhecimento.



Classificação**Processo Indutivo**

- A indução é a forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos.
- Na indução, um conceito é aprendido efetuando inferência indutiva sobre os exemplos apresentados. Portanto, as hipóteses geradas através da inferência indutiva podem ou não, dizer a verdade.



Classificação

Inferência Indutiva

- A inferência indutiva é um dos principais métodos utilizados para derivar conhecimento novo e prever eventos futuros.
- O aprendizado indutivo pode ser dividido em supervisionado e não-supervisionado.



Classificação**Hierarquia do Aprendizado Indutivo**

- No aprendizado não-supervisionado, o indutor analisa os exemplos e tenta determinar se alguns deles podem ser agrupados, formando clusters ou agrupamentos.
- No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido.



Árvores de Decisão

“
Árvore de decisão utiliza treinamento supervisionado para classificação e previsão de dados.

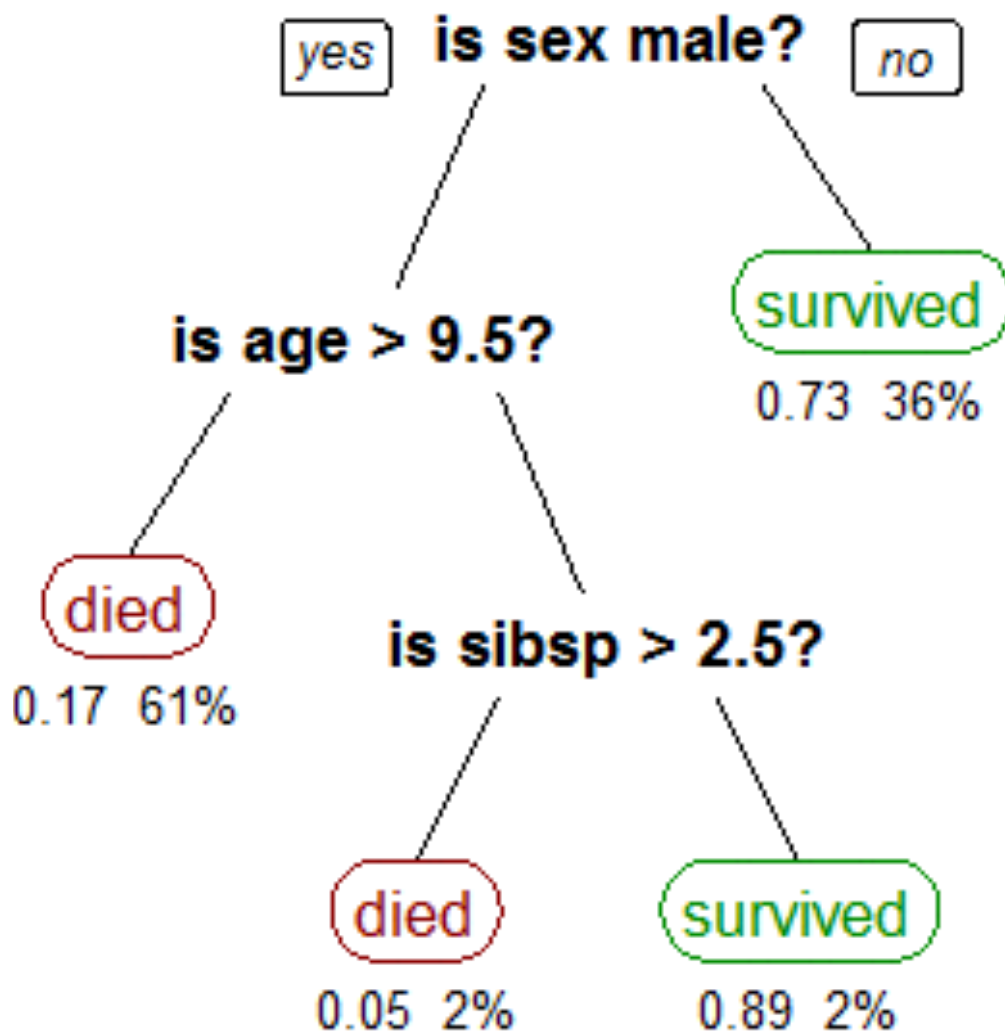


Árvores de Decisão

“
Árvore de decisão classificam instâncias ordenando-as da raiz para algum nó folha.



Árvores de Decisão



- Uma **árvore de decisão** é uma representação de uma tabela de decisão sob a forma de uma árvore.
- Tem a mesma utilidade da tabela de decisão.
- Trata-se de uma maneira alternativa de expressar as mesmas regras que são obtidas quando se constrói a tabela.



Árvores de Decisão

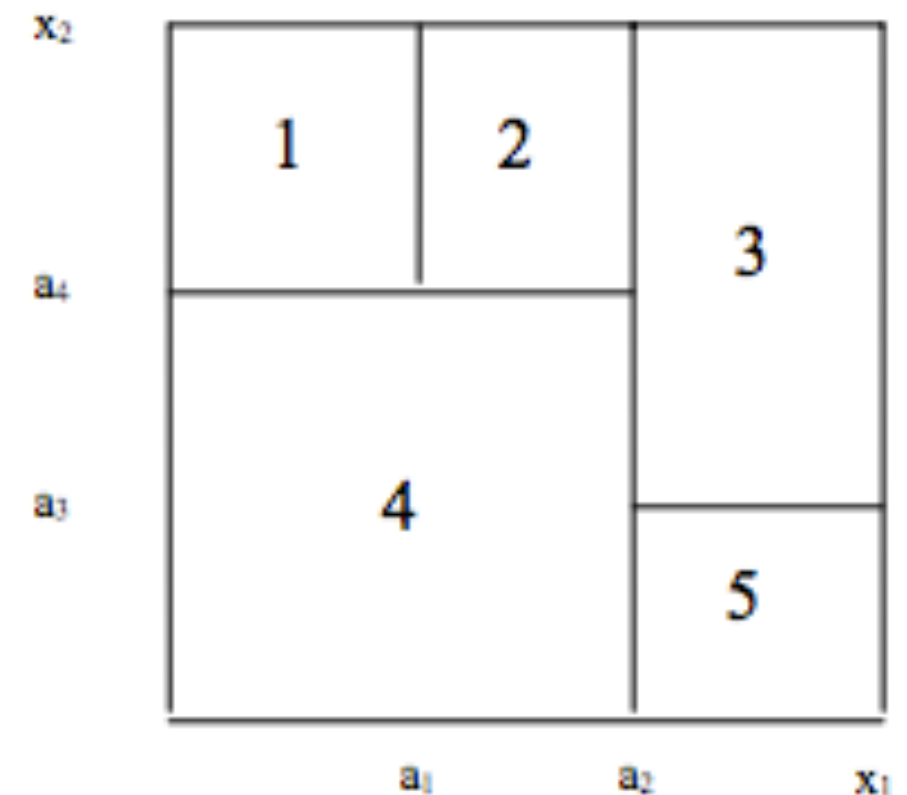
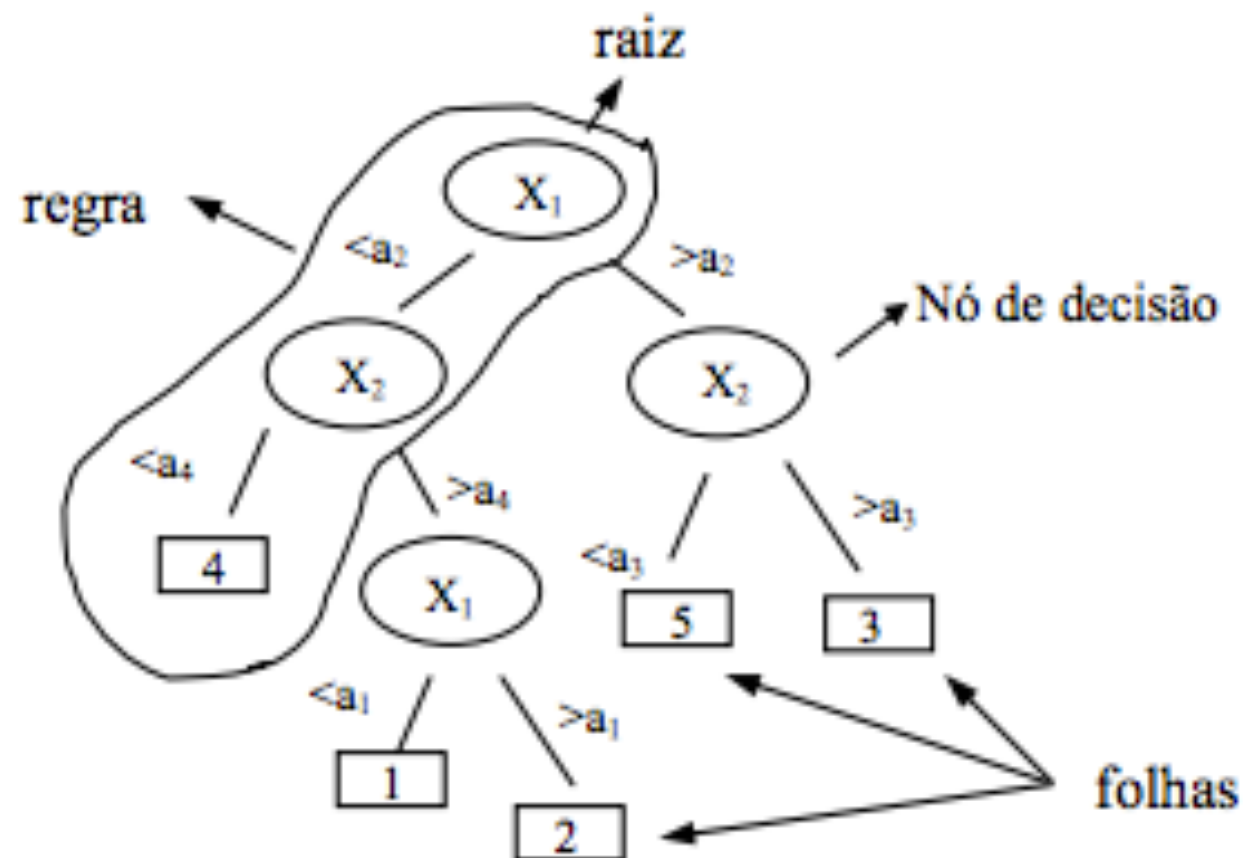
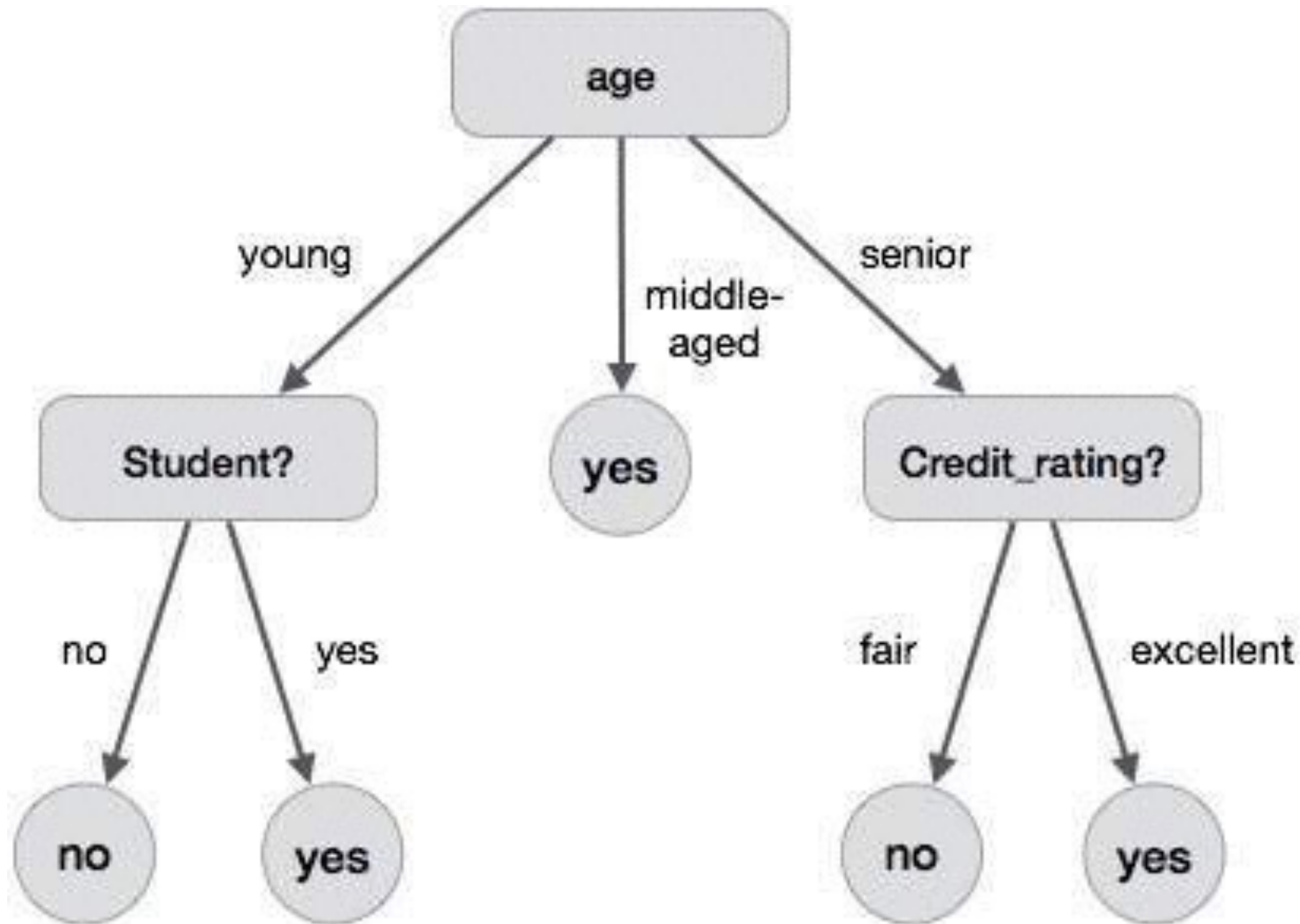


Figura 12 - Representação de uma árvore de decisão e sua respectiva representação no espaço

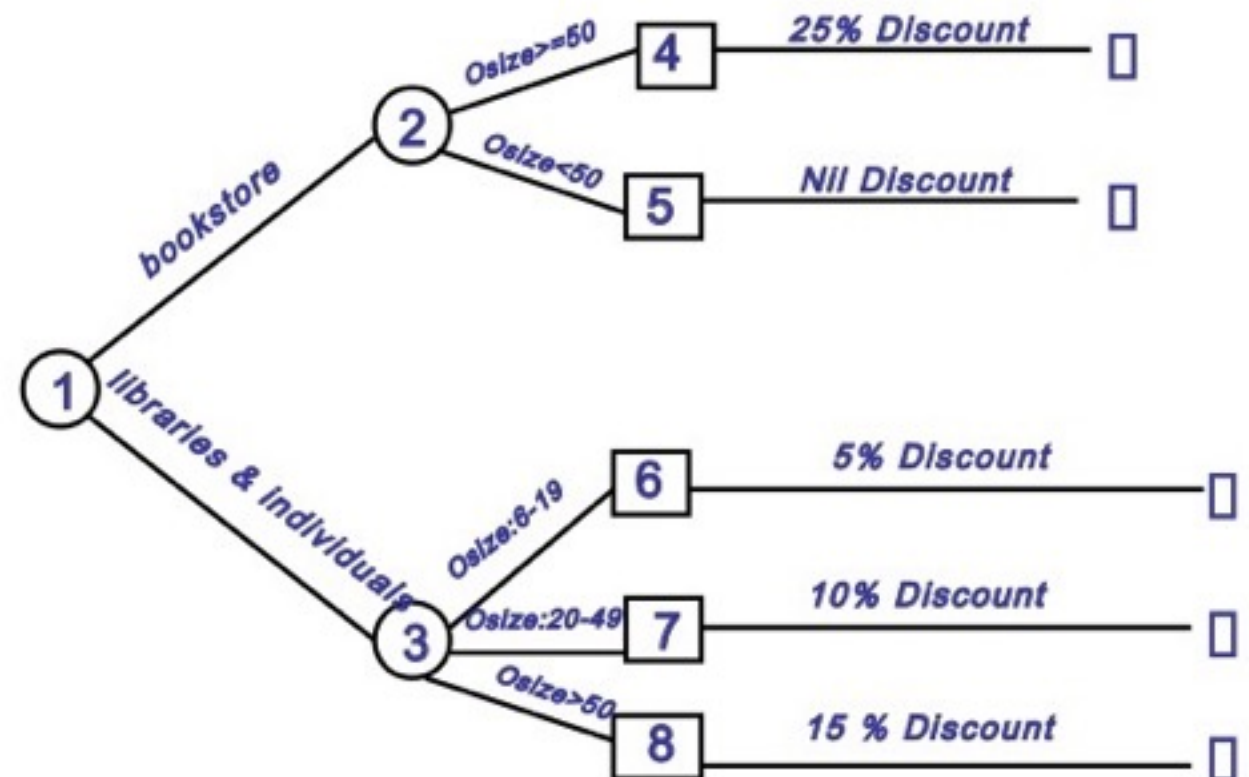
Fonte: Gama, 2004

Árvores de Decisão



Árvores de Decisão

- A árvore de decisão chega a sua decisão pela execução de uma seqüência de testes. Cada nó interno da árvore corresponde a um teste do valor de uma das propriedades, e os ramos deste nó são identificados com os possíveis valores de teste. Cada nó folha da árvore especifica o valor de retorno se a folha for atingida.



Árvores de Decisão

- Uma fórmula para calcular o quanto uma amostra é homogênea.
- Uma amostra completamente homogênea tem entropia zero.
- Uma amostra completamente heterogênea tem entropia 1.
- A fórmula da entropia é:

$$E(S) = \sum_{i=1}^c p_i \log_2(p_i)$$



Árvores de Decisão

- Para construir uma árvore de decisão, precisamos calcular dois tipos de entropia através de tabelas de frequência da seguinte forma:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Jogo	
Sim	Não
9	5
Prob = 9/14	5/14
$-(9/14) * \log(9/14; 2)$	$-(5/14) * \log(5/14; 2)$
Entropia dos dados (S1) = 0,0940	

Gini Index

Função de custo utilizada para separar os dados

```
1 proportion = count(class_value) / count(rows)
```

The proportions for this example would be:

```
1 group_1_class_0 = 2 / 2 = 1
```

```
2 group_1_class_1 = 0 / 2 = 0
```

```
3 group_2_class_0 = 0 / 2 = 0
```

```
4 group_2_class_1 = 2 / 2 = 1
```

Gini is then calculated for each child node as follows:

```
1 gini_index = sum(proportion * (1.0 - proportion))
```

```
2 gini_index = 1.0 - sum(proportion * proportion)
```

Gini Index

Função de custo utilizada para separar os dados

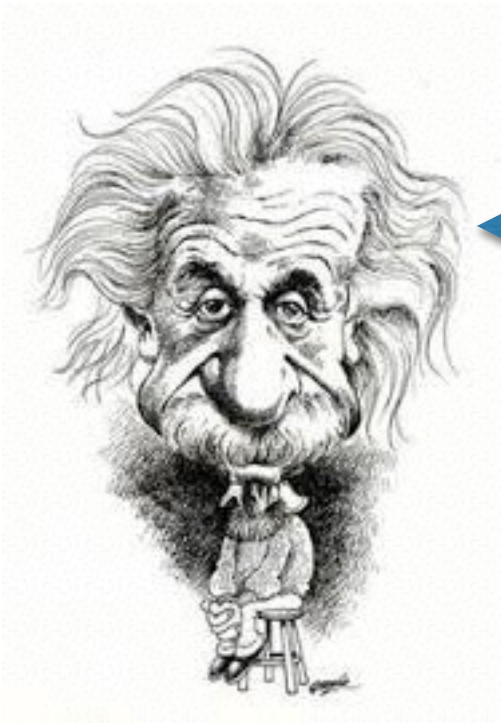
```
1 gini_index = (1.0 - sum(proportion * proportion)) * (group_size/total_samples)
```

In this example the Gini scores for each group are calculated as follows:

```
1 Gini(group_1) = (1 - (1*1 + 0*0)) * 2/4  
2 Gini(group_1) = 0.0 * 0.5  
3 Gini(group_1) = 0.0  
4 Gini(group_2) = (1 - (0*0 + 1*1)) * 2/4  
5 Gini(group_2) = 0.0 * 0.5  
6 Gini(group_2) = 0.0
```


Funções no Python

```
01 def hello(meu_nome, idade):  
02     print('Olá', meu_nome, '\nSua idade é:', idade)
```

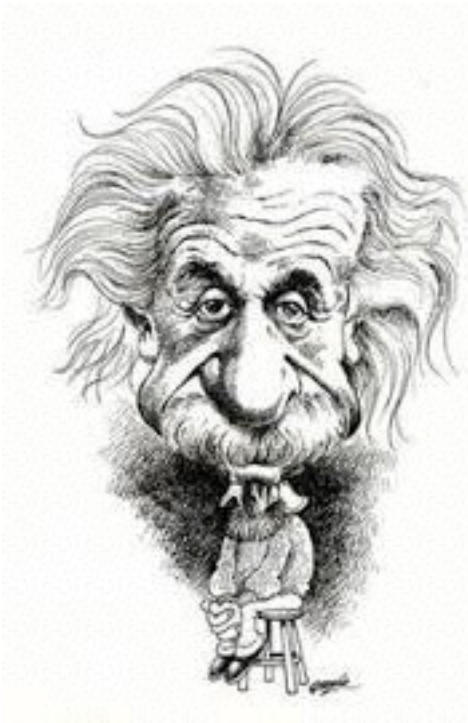


Vamos aprender a
criar funções
no python



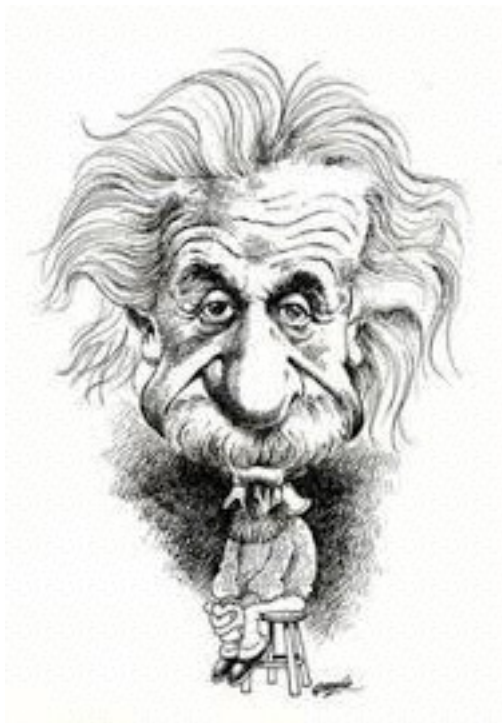
Funções no Python

```
01 def calcular_pagamento(qtd_horas,
valor_hora):
02     horas = float(qtd_horas)
03     taxa = float(valor_hora)
04     if horas <= 40:
05         salario=horas*taxa
06     else:
07         h_excd = horas - 40
08         salario = 40*taxa+(h_excd*(1.5*taxa))
09     return salario
```



Funções no Python

```
01 str_horas= input('Digite as horas: ')  
02 str_taxa=input('Digite a taxa: ')  
03 total_salario =  
   calcula_pagamento(str_horas,str_taxa)  
04 print('O valor de seus rendimentos é R  
$',total_salario)
```



Gini Index- Coding Dojo



```
1 # Calculate the Gini index for a split dataset
2 def gini_index(groups, classes):
3     # count all samples at split point
4     n_instances = float(sum([len(group) for group in groups]))
5     # sum weighted Gini index for each group
6     gini = 0.0
7     for group in groups:
8         size = float(len(group))
9         # avoid divide by zero
10        if size == 0:
11            continue
12        score = 0.0
13        # score the group based on the score for each class
14        for class_val in classes:
15            p = [row[-1] for row in group].count(class_val) / size
16            score += p * p
17        # weight the group score by its relative size
18        gini += (1.0 - score) * (size / n_instances)
19    return gini
20
21 # test Gini values
22 print(gini_index([[[1, 1], [1, 0]], [[1, 1], [1, 0]], [0, 1]]))
23 print(gini_index([[[1, 0], [1, 0]], [[1, 1], [1, 1]], [0, 1]]))
```

Separando os dados

```
2 def test_split(index, value, dataset):
3     left, right = list(), list()
4     for row in dataset:
5         if row[index] < value:
6             left.append(row)
7         else:
8             right.append(row)
9     return left, right
```



Definindo ponto de corte dos dados

```
1 # Select the best split point for a dataset
2 def get_split(dataset):
3     class_values = list(set(row[-1] for row in dataset))
4     b_index, b_value, b_score, b_groups = 999, 999, 999, None
5     for index in range(len(dataset[0])-1):
6         for row in dataset:
7             groups = test_split(index, row[index], dataset)
8             gini = gini_index(groups, class_values)
9             if gini < b_score:
10                 b_index, b_value, b_score, b_groups = index, row[index], gini, groups
11     return {'index':b_index, 'value':b_value, 'groups':b_groups}
```



Momento Pesquisa

1. ESPECIALIZAÇÃO:

De forma mais ampla, todo curso que é realizado após o ensino superior é chamado de curso de pós-graduação. Contudo, no país, o termo abreviado "pós" foi comumente associado a cursos de especialização. Essa modalidade de pós-graduação é vista por muitos candidatos como uma oportunidade de mudar de área. É comum os cursos de especialização estarem abertos a graduados de qualquer área de conhecimento. De acordo com o MEC, normalmente, a "pós" é um curso que tem o objetivo técnico ou profissional mais específico, sem abranger totalmente uma área de conhecimento. No entanto, ele precisa seguir algumas regras mínimas. Caso contrário, ele será considerado um simples curso livre. Conheça as principais exigências da especialização:

#Perfil do candidato: requisito mínimo exigido é que o candidato tenha diploma de curso superior

#Como escolher os cursos? Pela falta de fiscalização e controle na criação dos cursos de especialização pelo MEC, não existe um sistema oficial disponibilizando uma relação completa desses cursos. Assim, a melhor forma de escolher a pós é analisando algumas informações que podem melhor indicar o nível de qualidade da instituição.

Momento Pesquisa



2. MBA:

O MBA (da sigla em inglês, *Master Business Administration*) não é um curso de mestrado, como o nome sugere. Junto com a especialização, ele é um curso "lato sensu". Sendo assim, muitas de suas características são comuns com a "pós". De acordo com o próprio MEC, os MBA's "nada mais são do que cursos de especialização em nível de pós-graduação na área de Administração".

Muitos dos cursos oferecidos são voltados para o campo dos negócios e da gestão. No entanto, é possível encontrar MBA's de outras áreas, como comunicação e saúde.

#Perfil do candidato: independente do segmento do MBA, o curso em si é visto como uma oportunidade de realização e troca de contatos profissionais. Para fomentar esse *networking* é comum algumas instituições exigirem que o candidato tenha tido experiências de trabalho já consolidadas. O diploma da graduação também é obrigatório.

Momento Pesquisa

3. MESTRADO:

Para funcionar, os cursos "stricto sensu" como os mestrados precisam ser recomendados pela Capes e reconhecidos pelo MEC. É essa agência do Ministério da Educação que avalia o curso e atesta sua qualidade. As notas vão de 1 a 7 (nota máxima). Os cursos precisam ter, ao menos, a nota 3. Aqueles que possuem nota 5 já são considerados com "elevado padrão de qualidade". Mas, para ter essa nota, é preciso que tenham cursos de doutorado, além do mestrado. O máximo possível é a nota 7, que significa que o curso tem um "desempenho claramente destacado", afirma a Capes.

#Perfil do candidato: é importante que o candidato tenha interesse na realização de pesquisas científicas. Muitos dos mestrandos pretendem seguir a carreira acadêmica, ou seja, querem ser professores universitários.

#Como escolher os cursos? Os cursos recomendados pela Capes podem ser consultados pela internet. No portal da agência é possível conferir as notas dos programas e também relatórios detalhados sobre a qualidade do corpo docente, das instalações e da proposta curricular do mestrado.

Doutorado

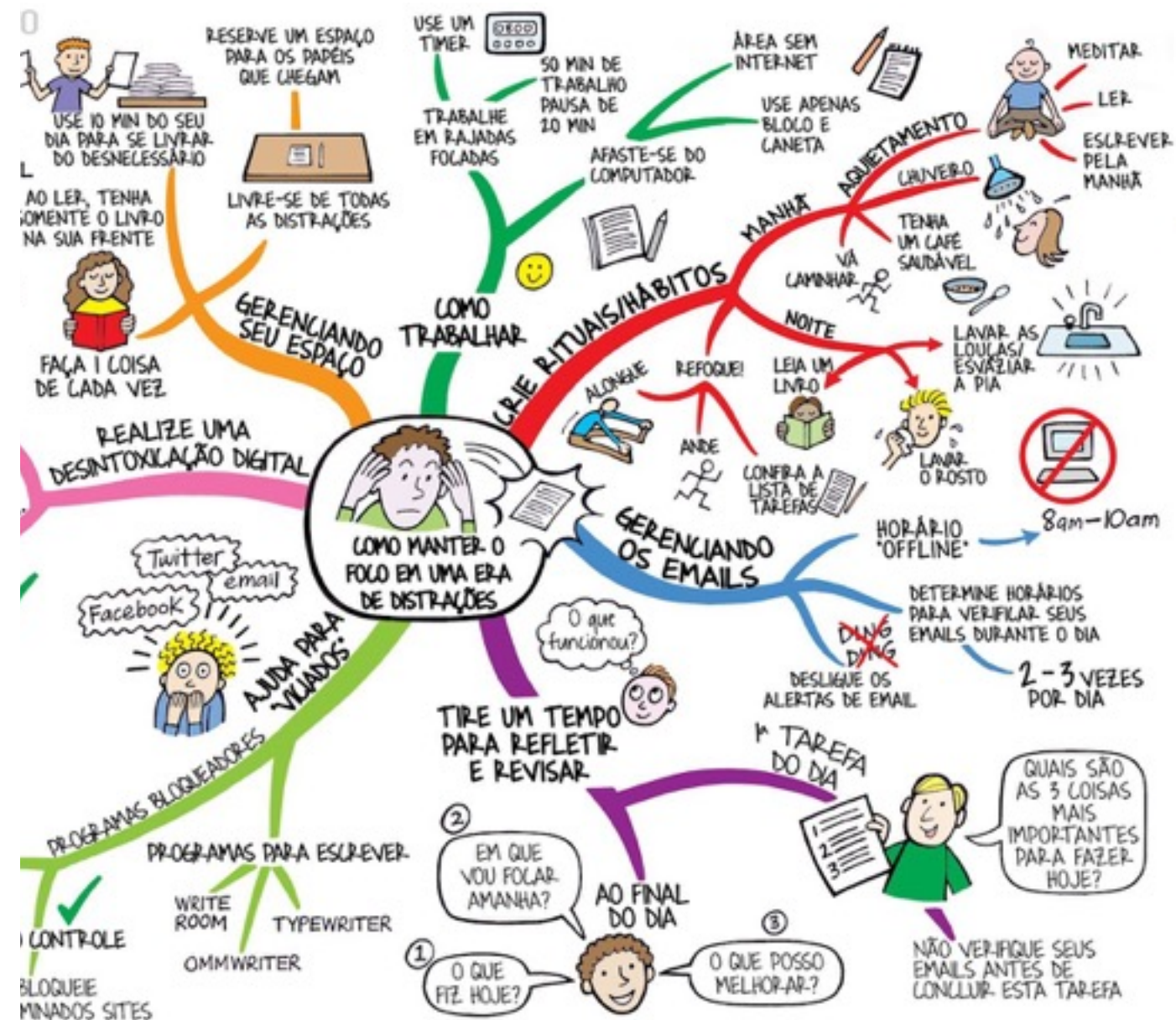
Uma das principais diferenças entre o mestrado e o doutorado está no nível de exigência que é solicitado aos candidatos selecionados no doutorado. Para ser professor de uma universidade pública, por exemplo, é preciso ter o doutorado.

Já depois dessa titulação, o candidato que finaliza o curso de doutorado ainda pode prosseguir nos estudos, realizando aprofundamentos de pesquisas. Tais atividades já são consideradas como de "pós-doc", ou seja, de pós-doutorado. No Brasil, ainda não existe a modalidade de doutorado profissional.

#Perfil do candidato: é recomendável que o postulante tenha forte interesse analítico no desenvolvimento de pesquisas científicas. É preferível que o candidato se dedique inteiramente ao doutorado em regime de dedicação exclusiva.

Mapa Mental

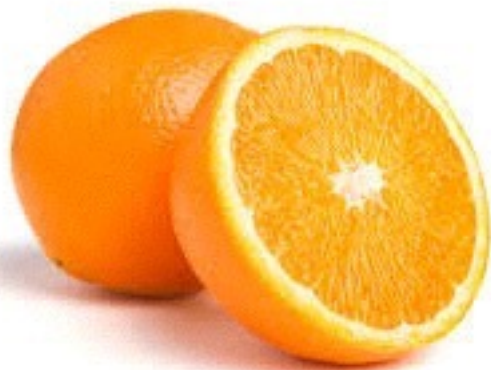
Mapa mental, ou mapa da mente é o nome dado para um tipo de diagrama, sistematizado pelo psicólogo inglês Tony Buzan, voltado para a gestão de informações, de conhecimento e de capital intelectual



Exercício

Exercício 1

Separem-se em equipe de até 5 pessoas e criem um mapa mental descrevendo o que foi feito na aula (10 minutos).



Ce sont les Batraciens, que le voyageur rencontre à chaque pas sur le sol japonais, ils fréquentent les lieux habités et peuplent les fertiles vallées, d'où une culture exultante à l'égard des animaux et les autres animaux sauvages. Fuyant la lumière et les lieux habités, la femelle des Batraciens à queue généralement dégingandée par le nom de Salamandre, se retire dans les profondeurs des montagnes riches en sources et en cascades. L'existence de différentes espèces de Salamandres au Japon, est d'une haute importance pour l'étude de la distribution géographique de ces reptiles, ainsi qu'elle l'a été jusqu'à ce jour l'Europe et l'Amérique du Nord pour leurs parents; mais ce qui est plus important encore c'est la découverte d'une espèce de Salamandre, qui par sa taille extraordinaire et sa forme spéciale, rappelle une création antédiluvienne, cette Salamandre le Géant des Batraciens est le représentant d'une race appartenant à cette longue période de notre globe qui sépare les formations houillères des terrains tertiaires, et qui vit apparues aux milieux de nos reptiles gigantesques et d'organisations bizarres. Je vais parler de l'époque débrisée, la célèbre Salamandre femelle des cantons d'Osaka qui depuis Sakaï jusqu'à Omiwa dont les dents ont été si vite l'objet de la mode japonaise, a été l'objet des spéculations des naturalistes.

Notre grande Salamandre (Salamandre japonaise) vit dans les profondes vallées des hautes montagnes de Nippon entre le 34° et 36° de lat. N.; elle séjourne dans les sources, dans les lacs et dans les lieux humides par les eaux pluviales ou même des cratères des Volcans éteints à une hauteur de 4 à 5000 pieds au dessus du niveau de la mer. Quelquefois elle quitte pendant la nuit les eaux qui lui servent d'habitat; mais son organisation et ses habitudes la rappellent bientôt dans cet élément, car elle meurt plus facilement que sur terre, une salamandre qui consiste en petits poissons, en grenouilles et en vers. C'est à Sakaï le petit village situé aux pieds du mont Sakaï-jima à 25 Mi environ à l'est de Mioko, que j'ai observé pour la première fois cette Salamandre. En de mes disciples, le Docteur Enryu, avait chargé un herboriste qui habite cette montagne de faire la recherche de ce rare et curieux animal. Malgré les renseignements des montagnards le Docteur Enryu n'eut rien de mieux à me rapporter de la grande Salamandre — ce n'est le plus souvent dans les montagnes d'Osaka-jima. Fui en le loupant des rapporter une vivante en Europe. Elle existe encore au Mont des Pays-Bas, on l'a trouvée une longueur d'environ trois pieds, taille extraordinaire que je n'ai jamais observée ailleurs au Japon. Cet individu a été depuis plusieurs années l'objet des observations de M. Schlegel, qui en a donné une description complète sous tous les rapports.



Francisco Nauber Bernardo Gois
Email: naubergois@gmail.com