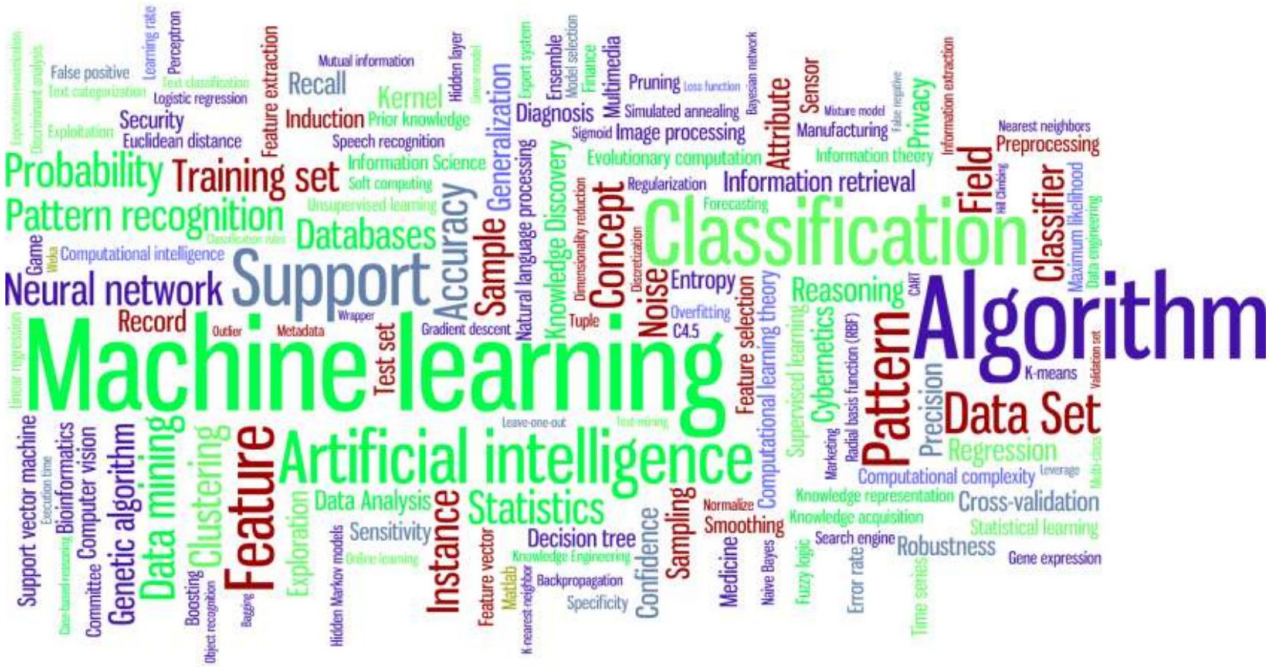
Universidade Federal do Paraná (UFPR)
Especialização em Engenharia Industrial 4.0

Uma introdução à *Machine Learning*



David Menotti
www.inf.ufpr.br/menotti/am-191

Hoje

- Introdução sobre AM (01-02)
- Agenda
  - Como aplicar ML na prática? (03-03)
  - Laboratório Weka: A ferramenta (04-04)
  - Artificial Intelligence & Machine Learning (05-05)
  - Deep Learning & i4.0 (06-06)
  - Laboratório Weka: Pré-processamento (07-08)
  - Classificadores & Lab (09-12)
  - Regressão & Lab (13-14)
  - *Clustering* & Lab (15-16)
  - + Classificadores (17-18) & Lab (19-20)

Aprendizado de Máquinas

- **Herbert Alexander Simon**:
  "Aprendizado é qualquer processo pelo qual um sistema melhora sua *performance* pela experiência."

- "Machine Learning está preocupado com programas de computador que automaticamente melhoram sua *performance* pela experiência. "

- Economista / Matemática 1913 - 2001
  - Simulaçao computacional da Cognição humana (~1954)

Herbert Simon
Turing Award 1975
Nobel Prize in Economics 1978
1913 - 2001

Por que *Machine Learning*?

- Desenvolver sistemas que podem automaticamente se adaptar e se customizar para usuários individuais.
  - Notícias personalizadas OU Filtro de email

- Descobrir novo conhecimento a partir / usando grandes bases de dados (*data mining*).
  - Análise de carrinho de supermer. (e.g. fraldas e cervejas)

Por que *Machine Learning*?

- Habilidade de imitar humanos e substituí-los em certas tarefas monótonas - que exigem alguma inteligência.
  - Como o reconhecimento de caracteres manuscritos

- Desenvolver sistemas que são muito difíceis / caros para construir manualmente porque eles requerem habilidades ou conhecimento detalhados específicos ajustados para uma tarefa específica (gargalo de engenharia do conhecimento).

Por que *AGORA*?

- Inundação de dados disponíveis
  - especialmente com o advento da internet.
- Incremento de força computacional
- Progresso crescente de:
  - algoritmos disponíveis
  - teoria desenvolvida por pesquisadores
- Aumento no suporte/apoio das indústrias

Aplicações em ML

O Conceito de Aprendizado

- **Aprendizado** = **Melhoria** com **experiência** em alguma **tarefa**
  - Melhoria sobre a tarefa **T**
  - Com respeito a medida de desempenho **D**
  - Baseado na experiência **E**

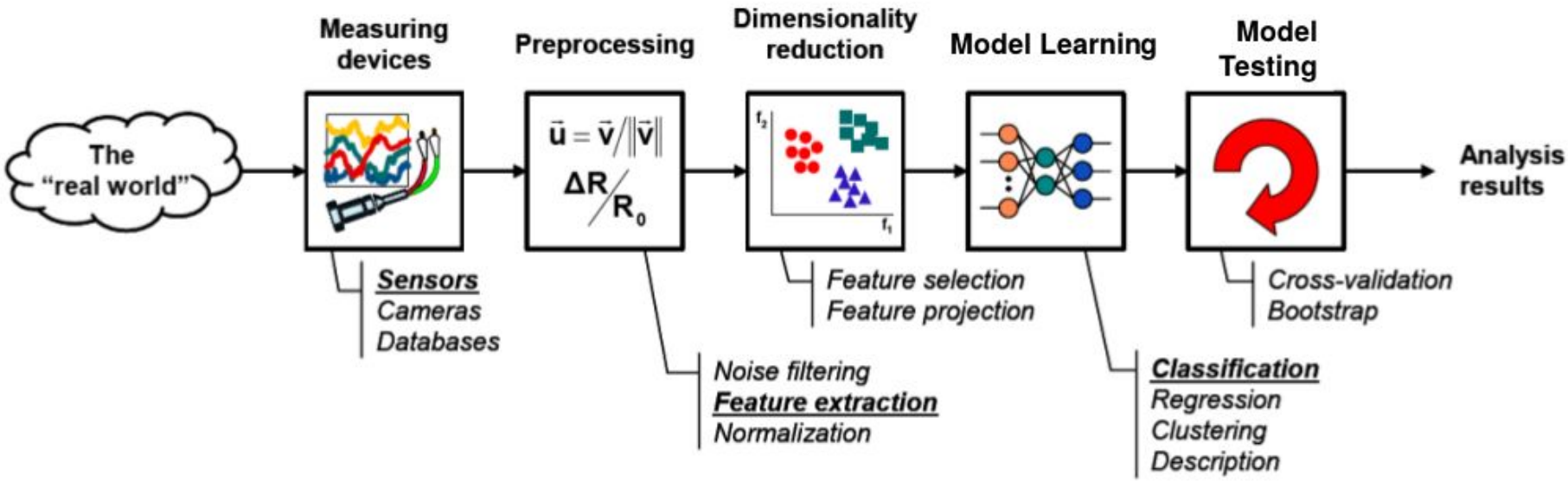
Motivação - Filtro de SPAM

- **Example**: *Spam Filtering*
  Spam (*Sending and Posting Advertisement in Mass*)
  - é todo email que o usuário não queria receber e não autorizou o recebimento

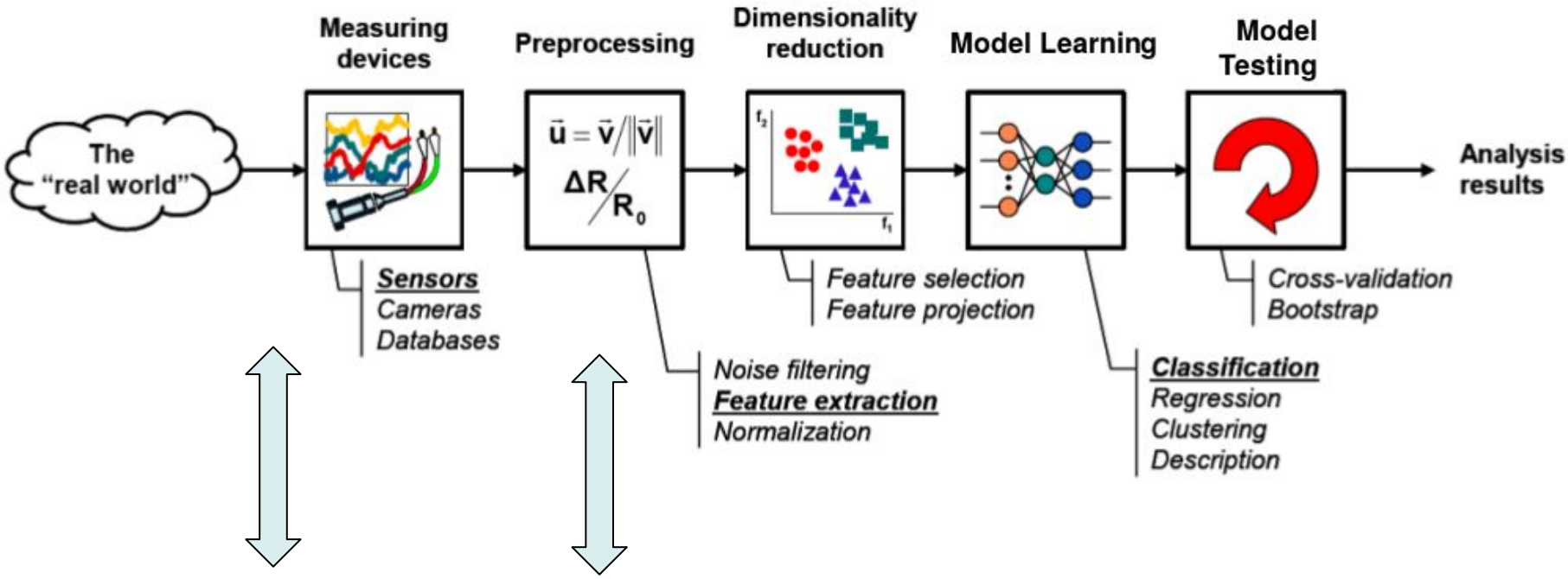
  T: Identificar emails SPAM
  D:
  % de emails spam que foram filtrados
  % de emails non-spam (ham) que foram incorretamente filtrados
  E: uma base de dados de emails que foram rotulados pelos usuários

O Processo de Aprendizado

O Processo de Aprendizado



O Processo de Aprendizado



Servidor de Emails

- Número de destinatários
- Tamanho da Mensagem
- Número de anexos
- Número de "re's" no assunto

A Base de Dados (*Data Set*)

Atributos

Atributo Meta

Instâncias

Número de novos destinatários	Tamanho do Email (kb)	País (IP)	Tipo de Cliente	Tipo de Email
0	2	Brasil	Ouro	Ok
1	4	Brasil	Prata	Ok
5	2	Argentina	Bronze	Spam
2	4	Russia	Bronze	Spam
3	4	Brasil	Bronze	Ok
0	1	EUA	Prata	Ok
4	2	EUA	Prata	Spam

Numéricos

Nominal

Ordinal

Aprendizado do Modelo

Base de Dados

Conjunto de Aprendizado (Treinamento)

Indutor

Algoritmo de Indução

Classificador

Modelo de Classificação

Avaliação do Modelo

Base de Dados

Conjunto de Avaliação (Teste)

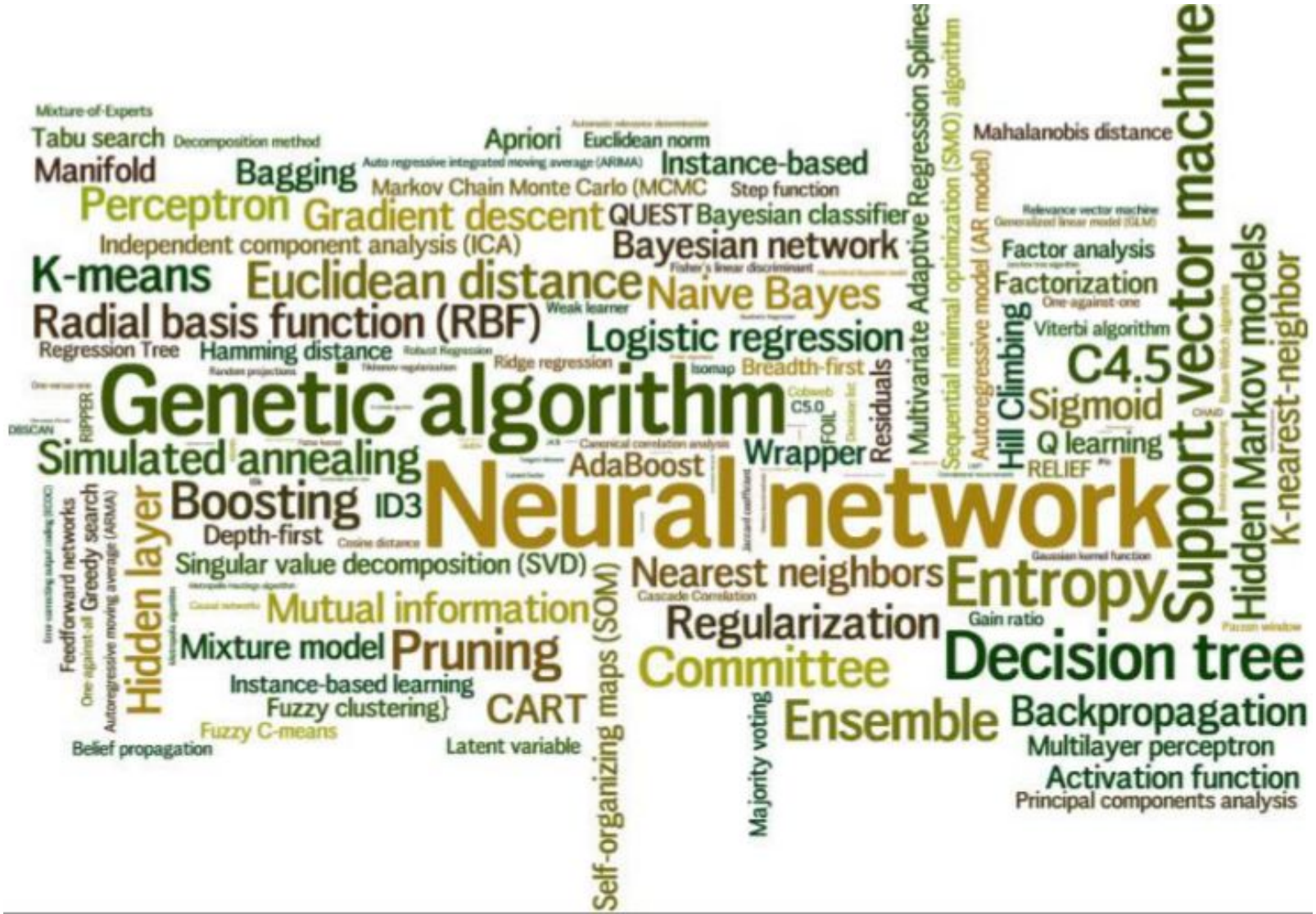
Indutor

Algoritmo de Indução

Classificador

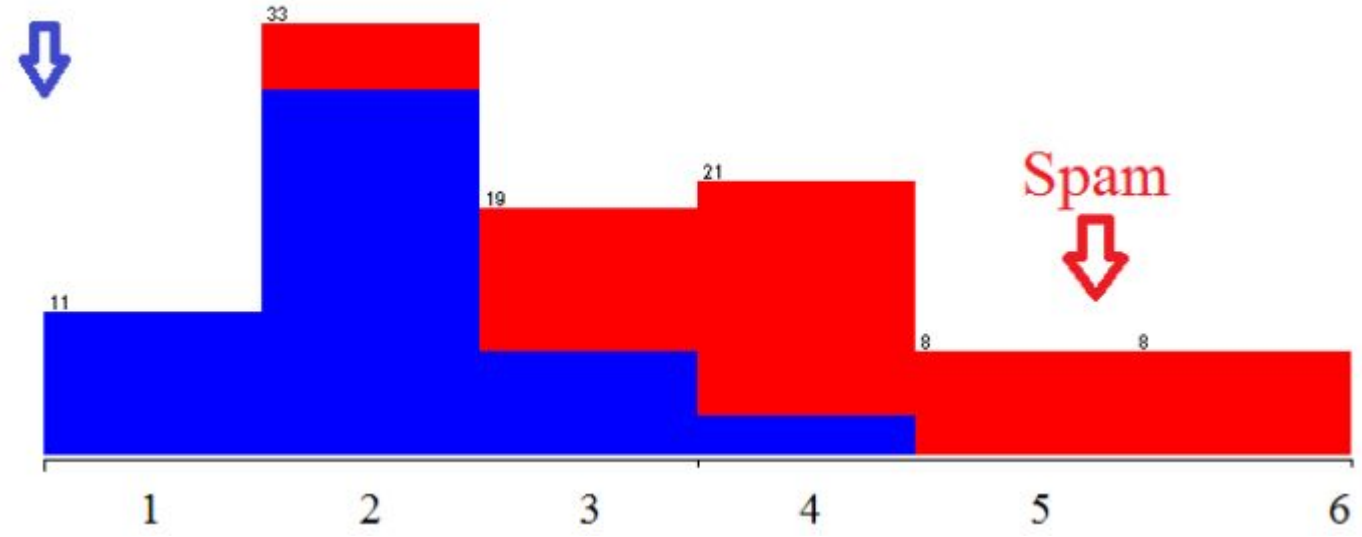
Modelo de Classificação

Algoritmos de Aprendizado



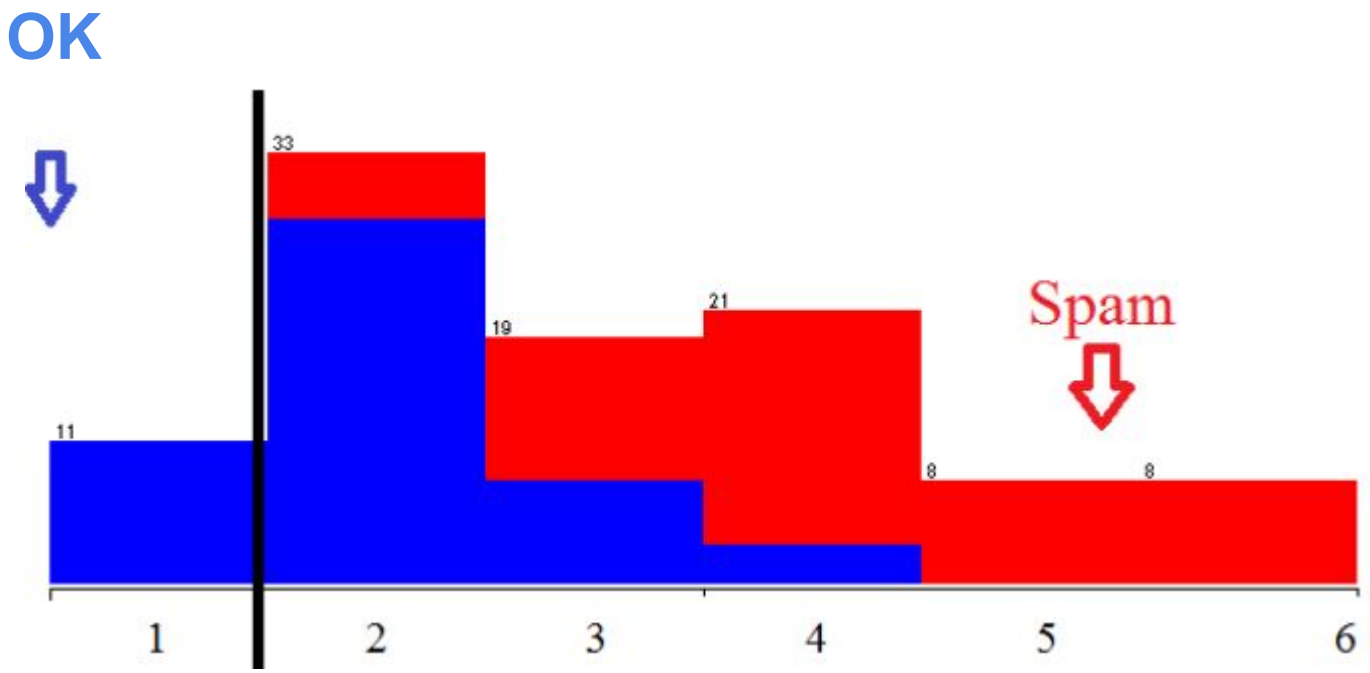
Análise da Classificação

OK

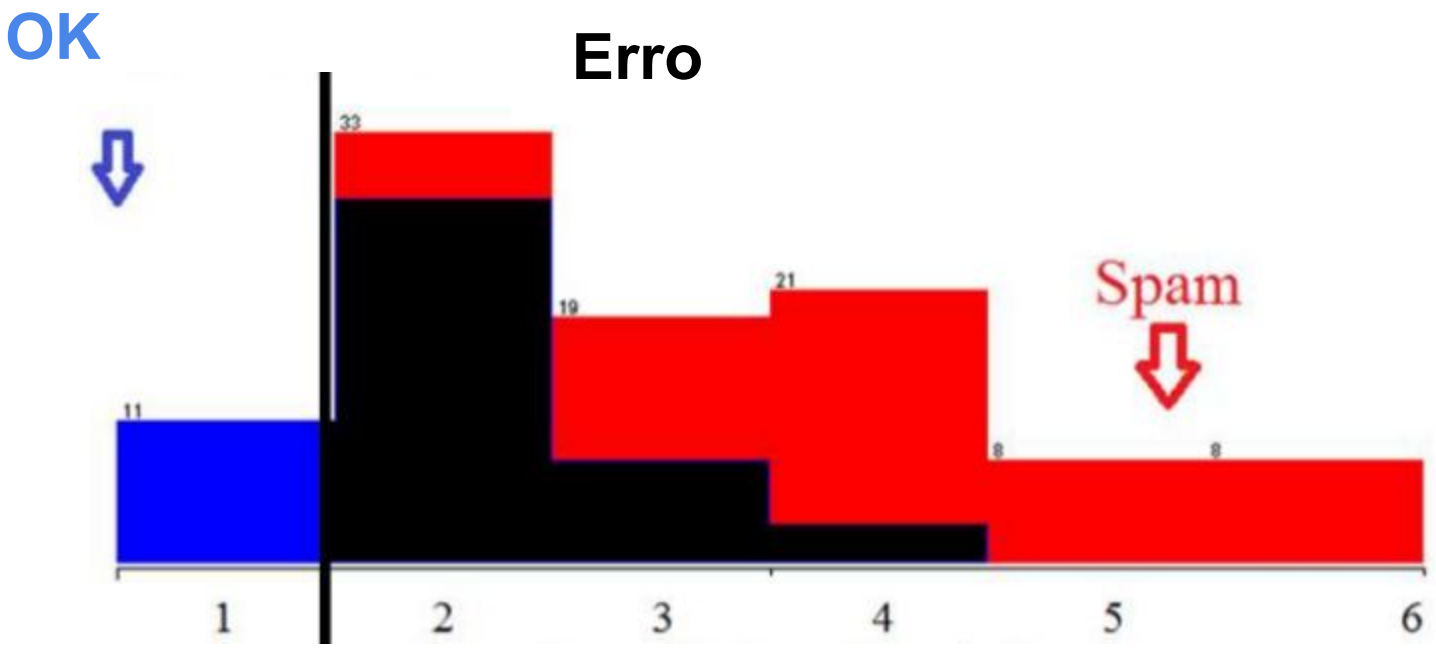


Número de novos destinatários

Análise da Classificação

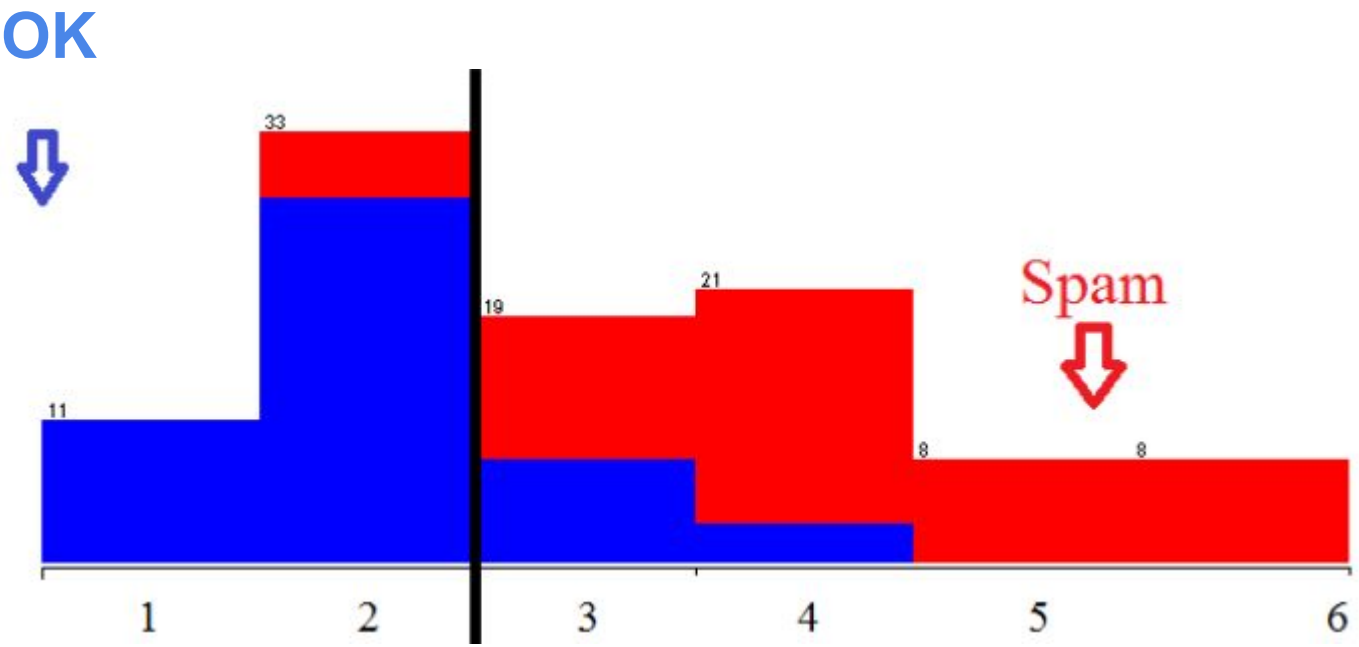


Análise da Classificação



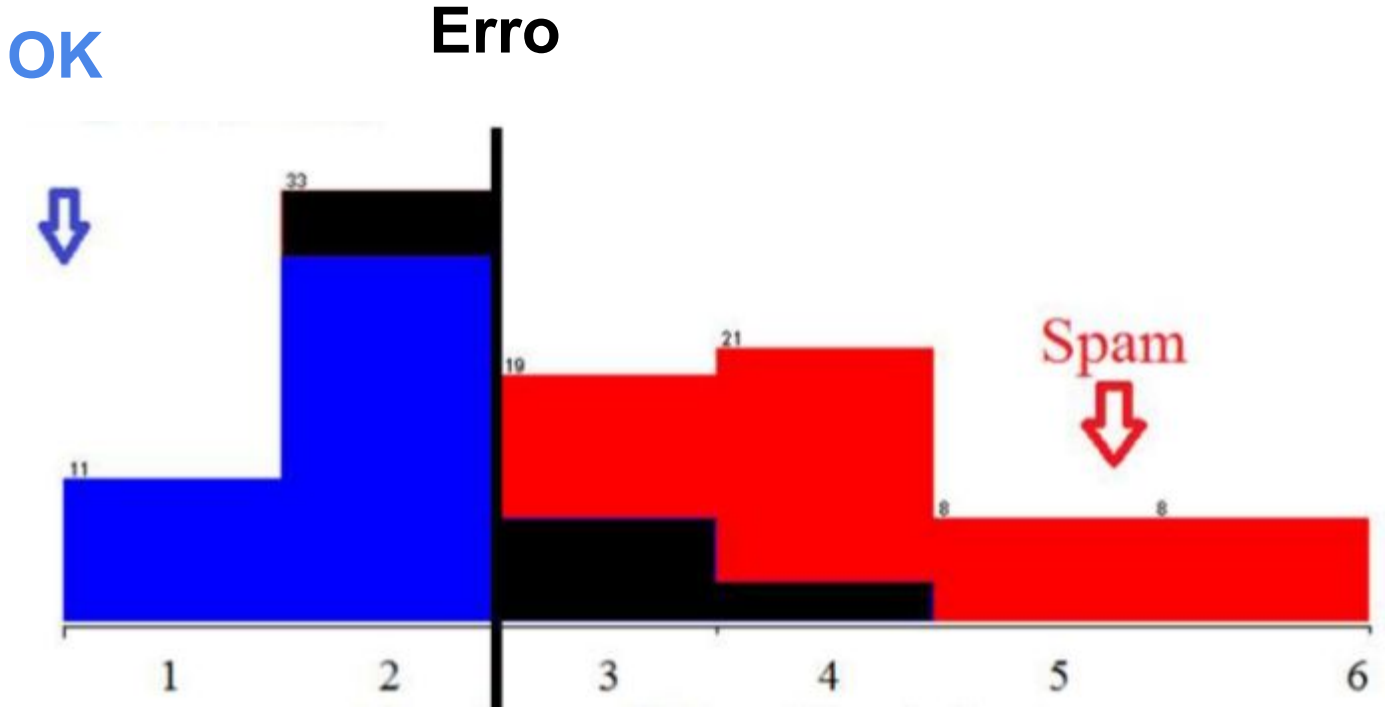
Número de novos destinatários

Análise da Classificação



Número de novos destinatários

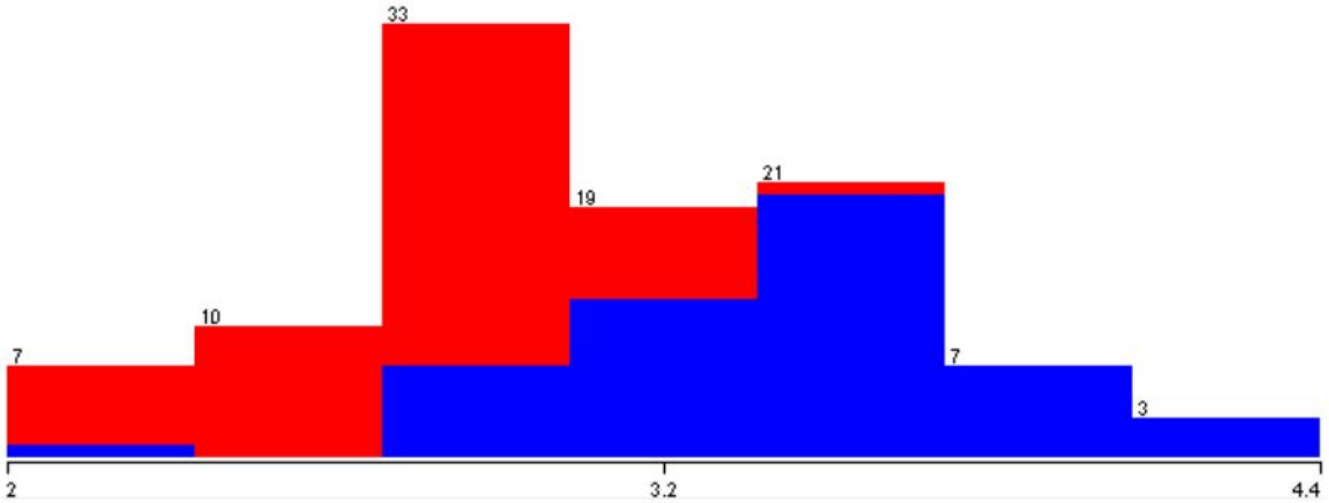
Análise da Classificação



Número de novos destinatários

Análise da Classificação

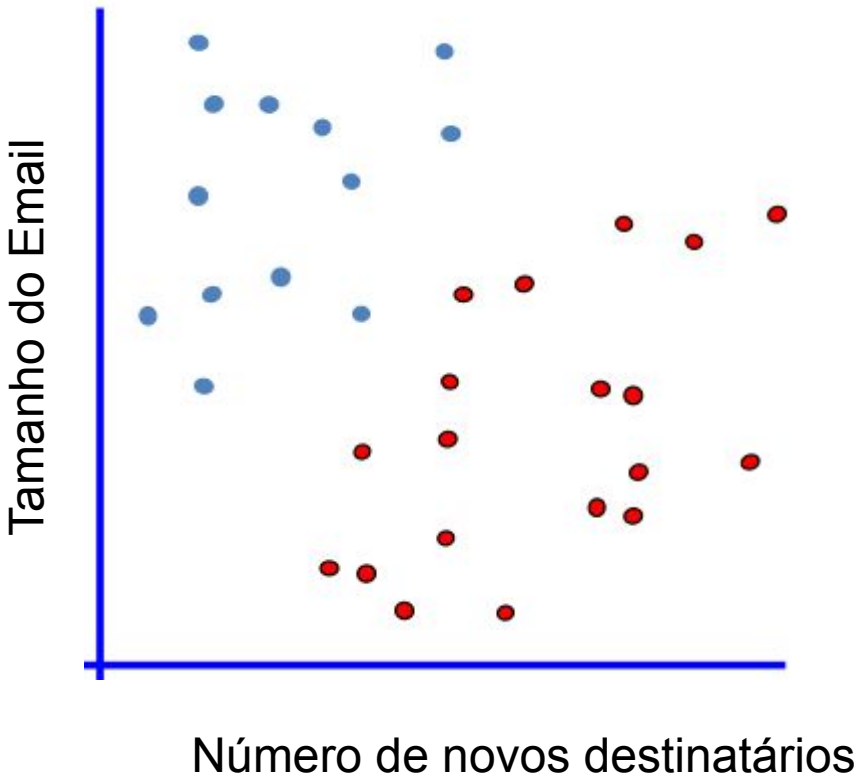
- Outro **atributo** - confusão?



Tamanho do Email (kbytes)

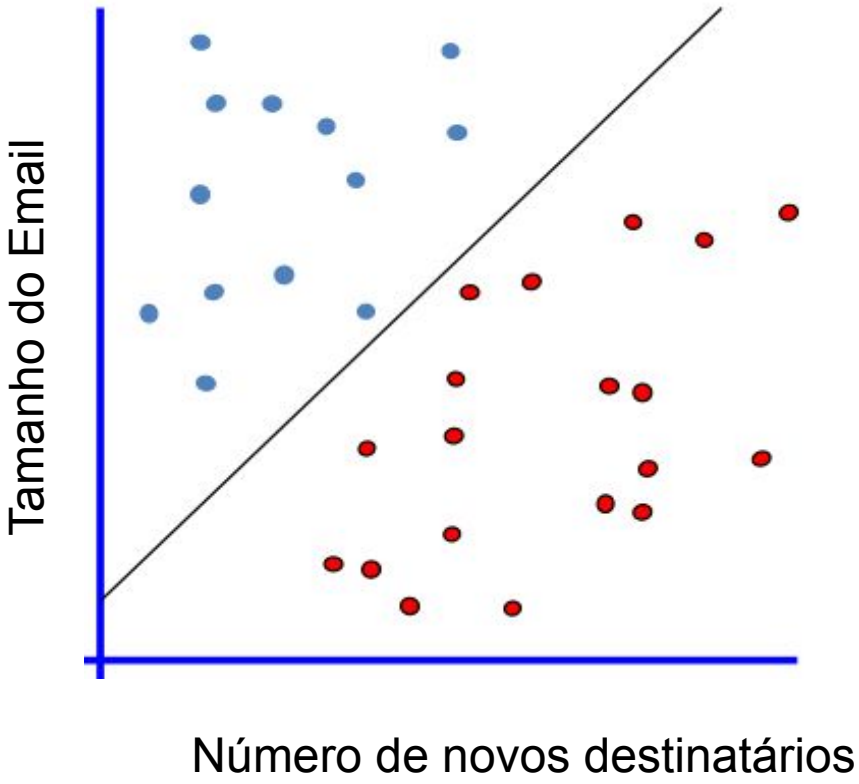
Classificadores Lineares

- Como você classificaria estes dados?



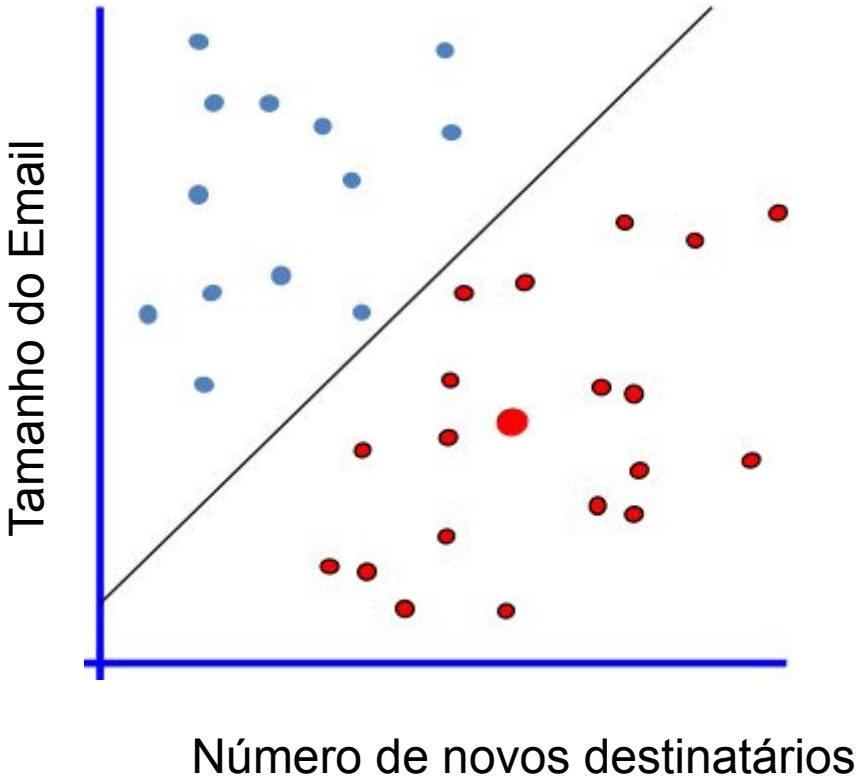
Classificadores Lineares

- Como você classificaria estes dados?



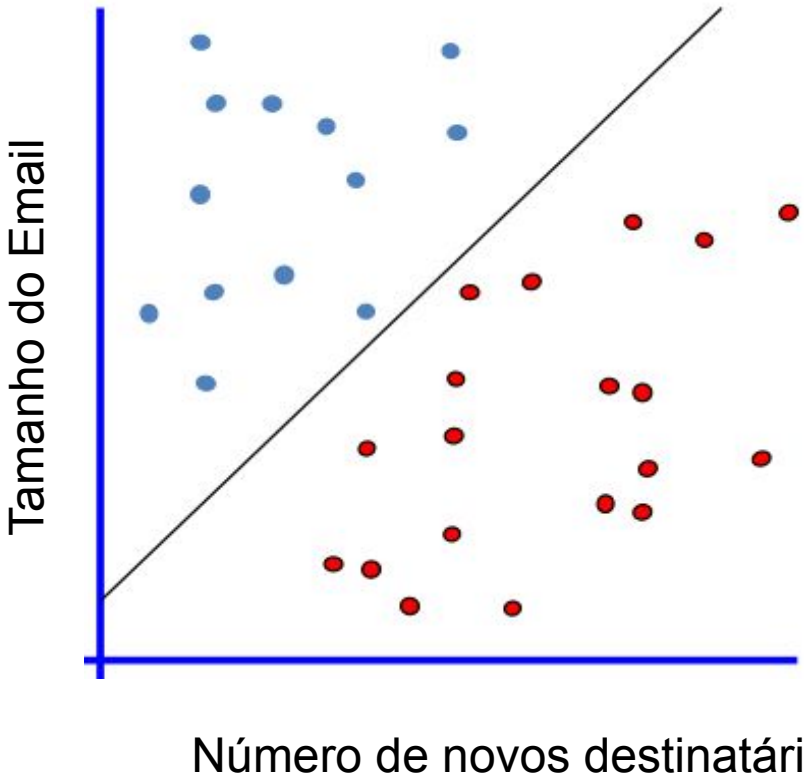
Quando um novo email é enviado

1. Coloca-se o novo email no espaço
2. Classifica-o de acordo com o subespaço no qual ele “reside”



Classificadores Lineares

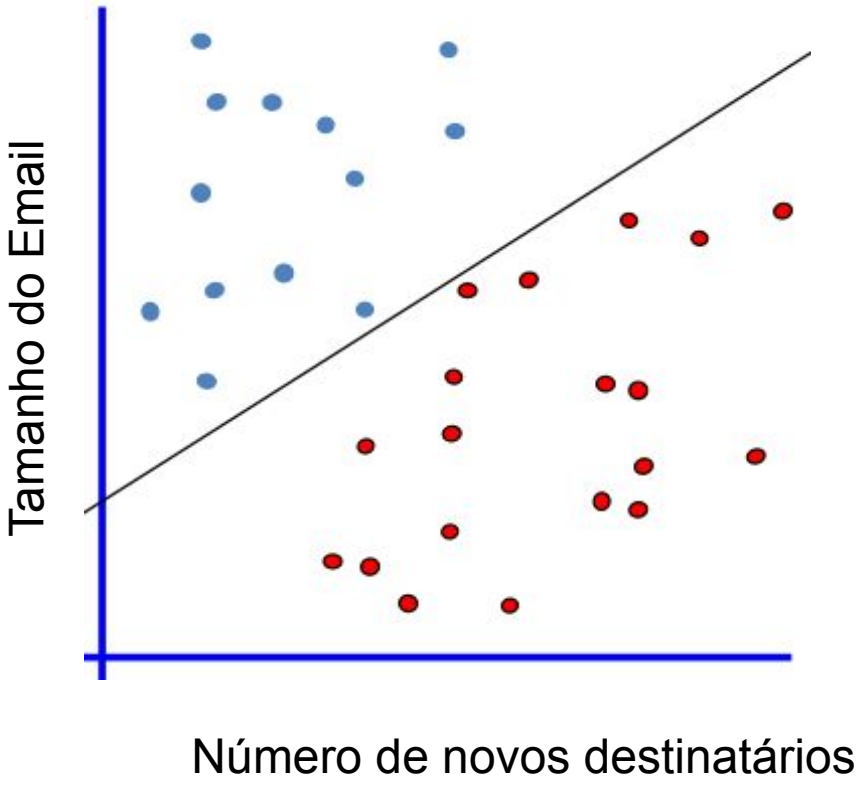
Como você classificaria estes dados?



- Várias separações são possíveis

Classificadores Lineares

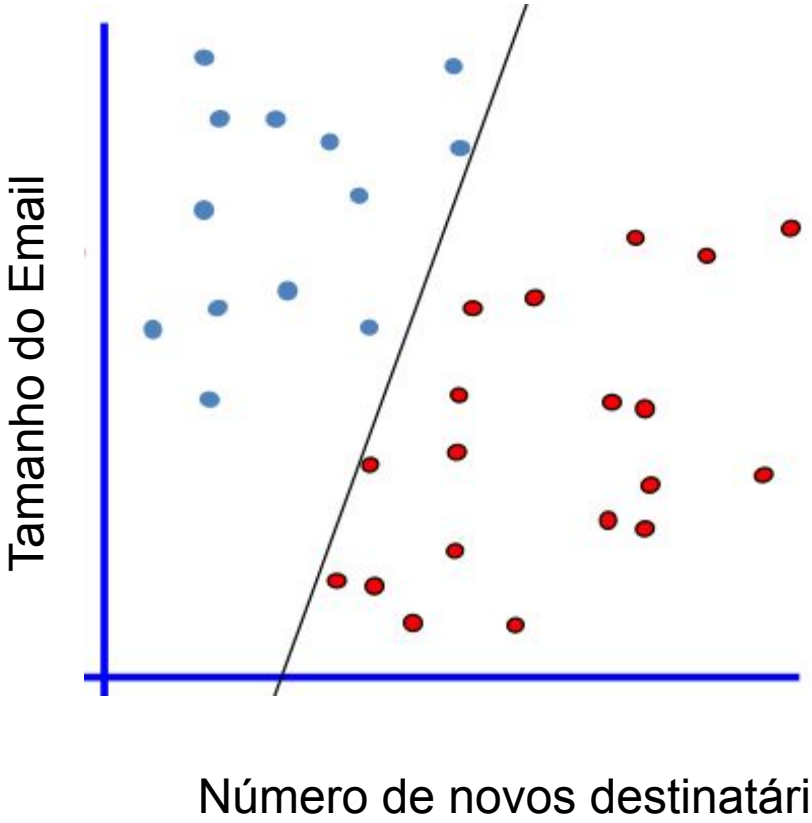
Como você classificaria estes dados?



- Várias separações são possíveis - Tamanho do Email?

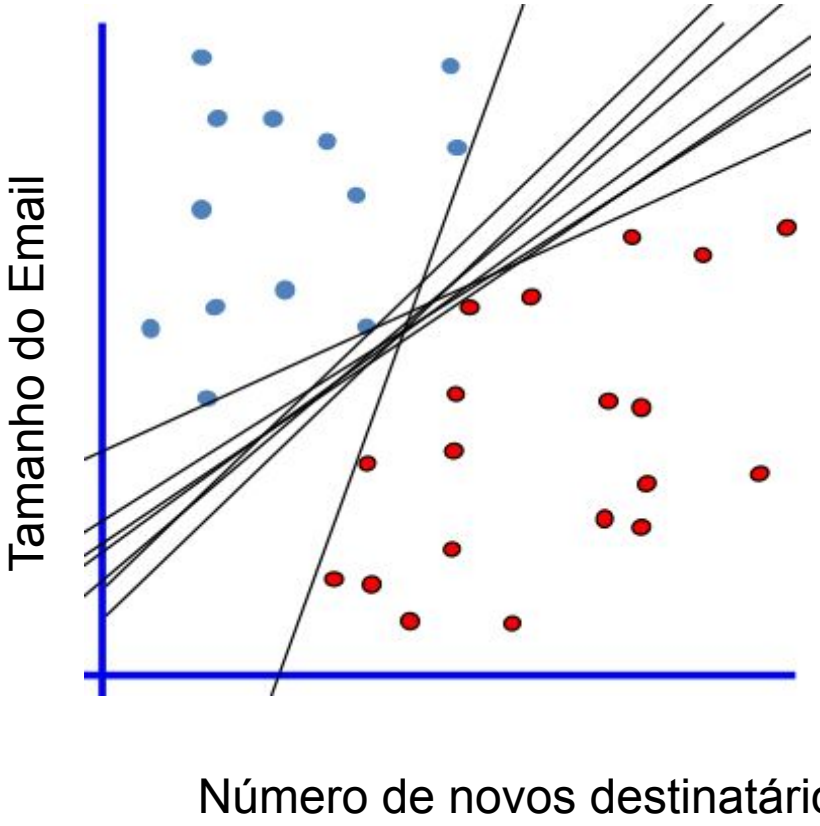
Classificadores Lineares

Como você classificaria estes dados?



- Várias separações são possíveis - Número de novos dest.?

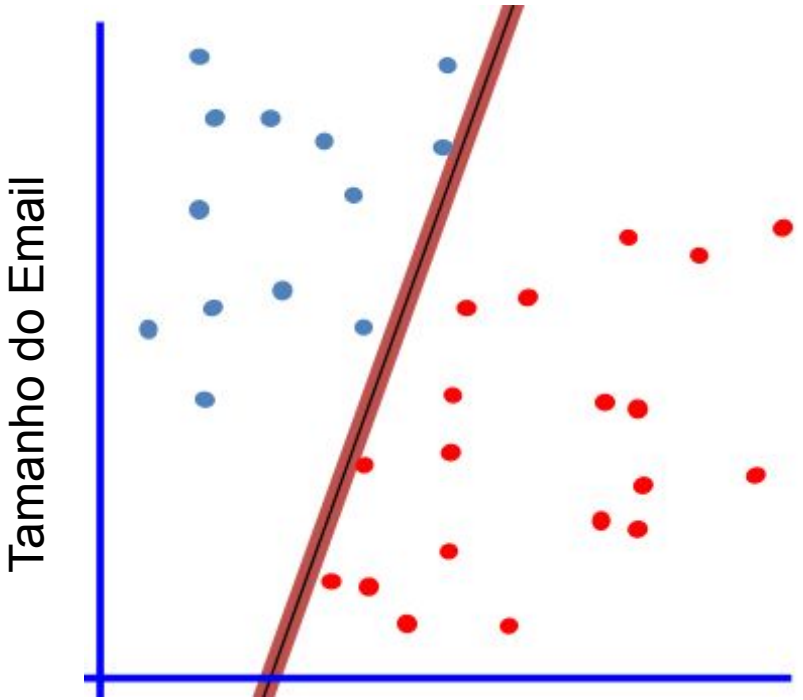
Classificadores Lineares



Qualquer uma delas seria uma boa escolha …

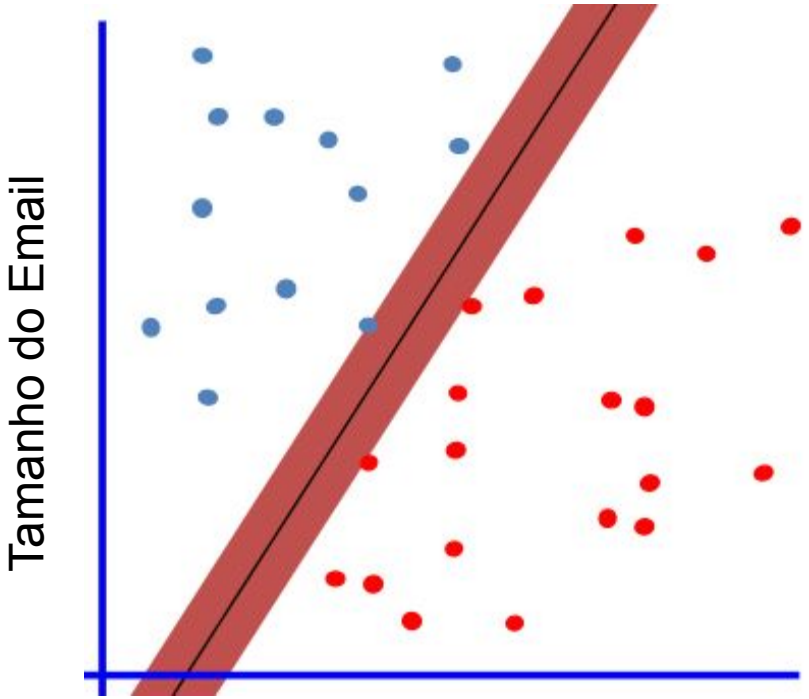
… mas qual é a melhor?

Margem Classificadora



Definir a margem de um classificador linear como a **largura** que o limite da margem pode ser aumentado antes de atingir/acertar um ponto

Margem Máxima



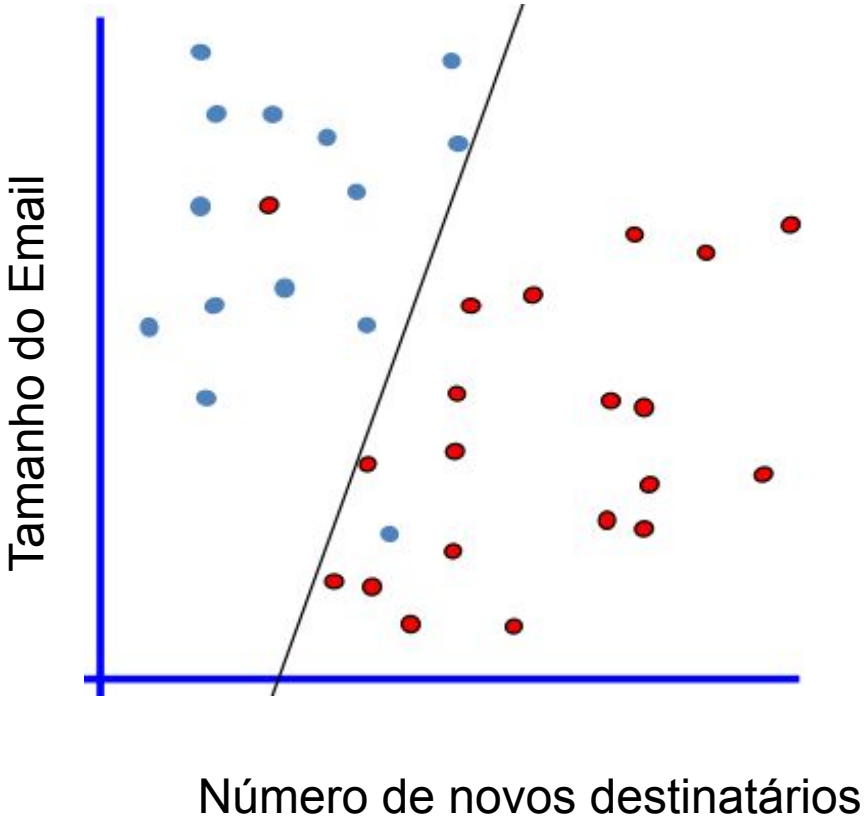
A margem máxima do classificador linear é o classificador linear com a margem máxima.

Este é o classificador mais simples do tipo SVM (Support Vector Machines) chamado de **LSVM**

Linear SVM

Nenhum Classificador Linear pode cobrir todas as instâncias

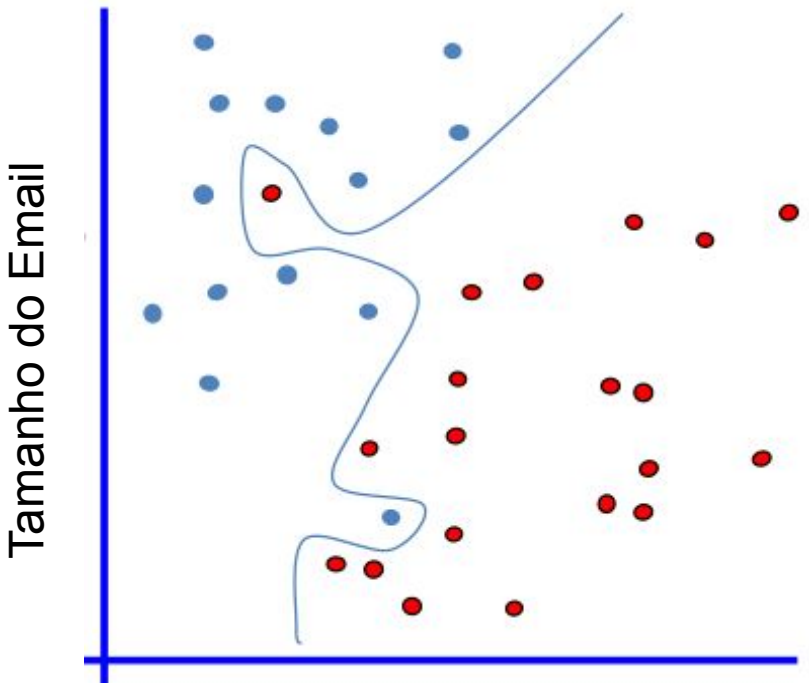
Como você classificaria estes dados?



Nenhum Classificador Linear

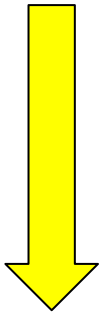
- Idealmente, a melhor fronteira de decisão deveria ser aquela que provê um desempenho ótimo tal como ...

Nenhum Classificador Linear pode cobrir todas as instâncias



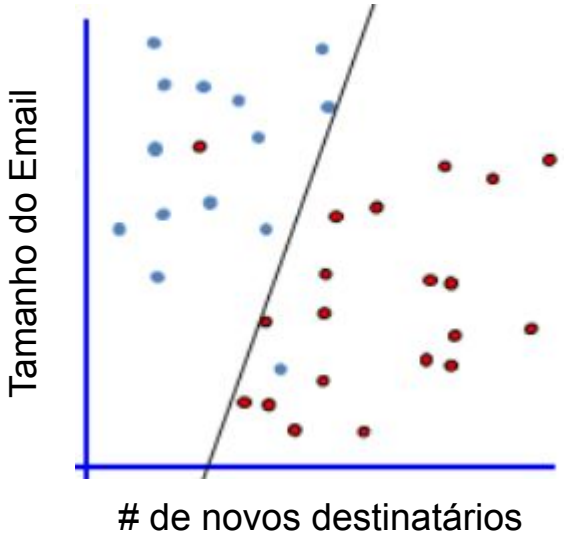
Nenhum Classificador Linear

- No evento, a satisfação imediata é prematura porque o objetivo central do desenvolvimento de um classificador é o de classificar corretamente uma nova entrada

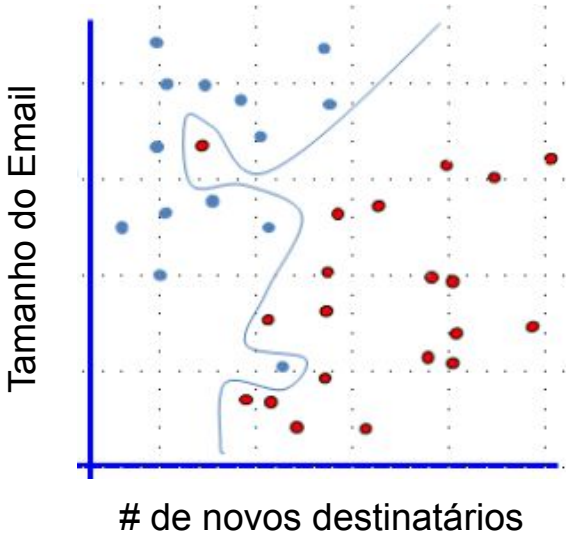


Problema de Generalização

Qual delas?



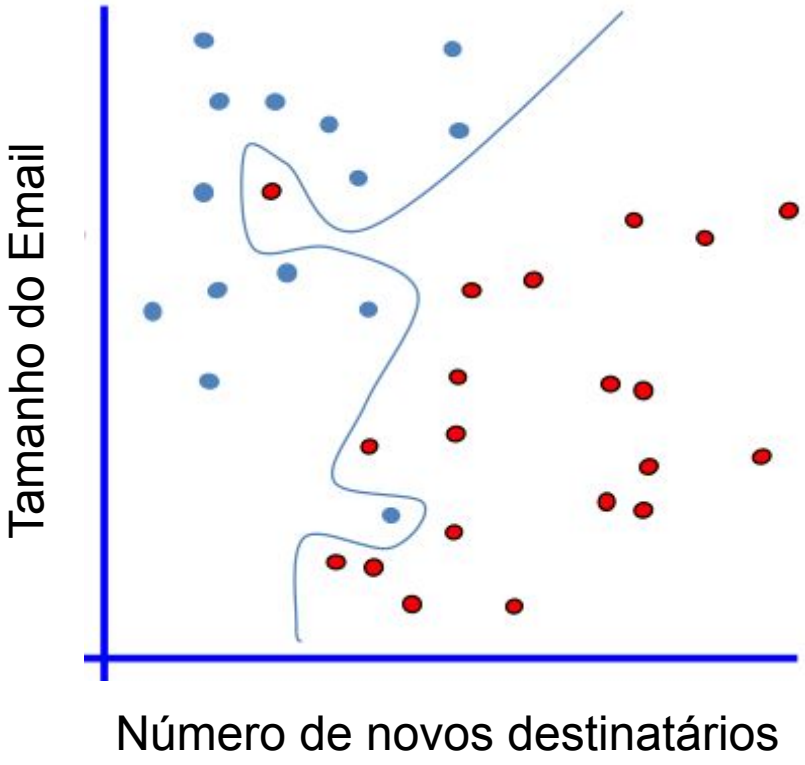
2 Erros
Modelo Simples



0 Erro
Modelo Complexo

Avaliando o que foi Aprendido

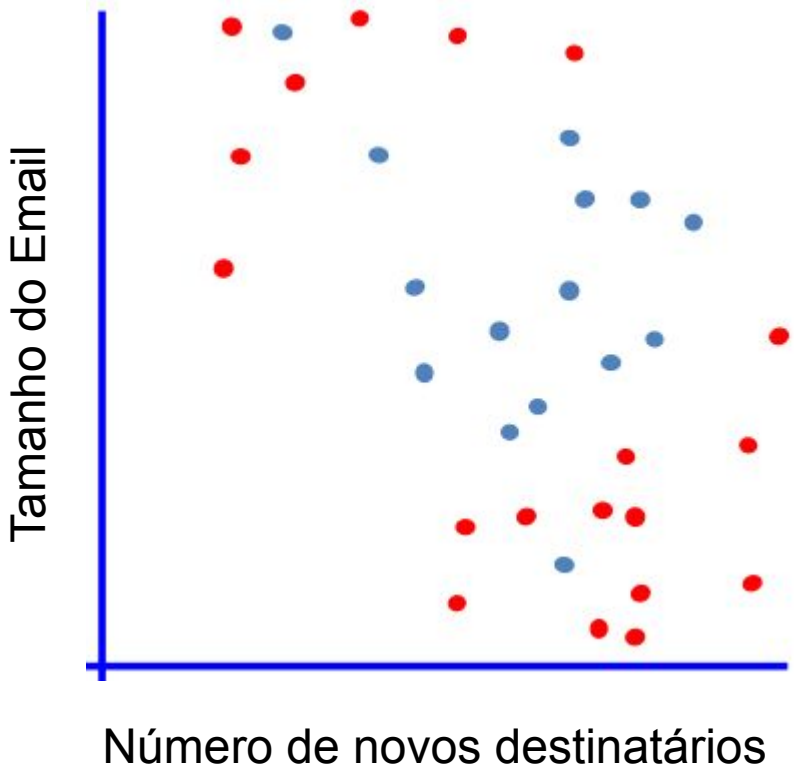
1. Aleatoriamente seleciona-se uma porção dos dados para ser usada para aprendizado (o conjunto de treinamento)
2. Aprende-se o modelo a partir do conjunto de treinamento
3. Uma vez treinado, o modelo é executado sobre as instâncias remanescentes (o conjunto de teste) para ver como ele se comporta



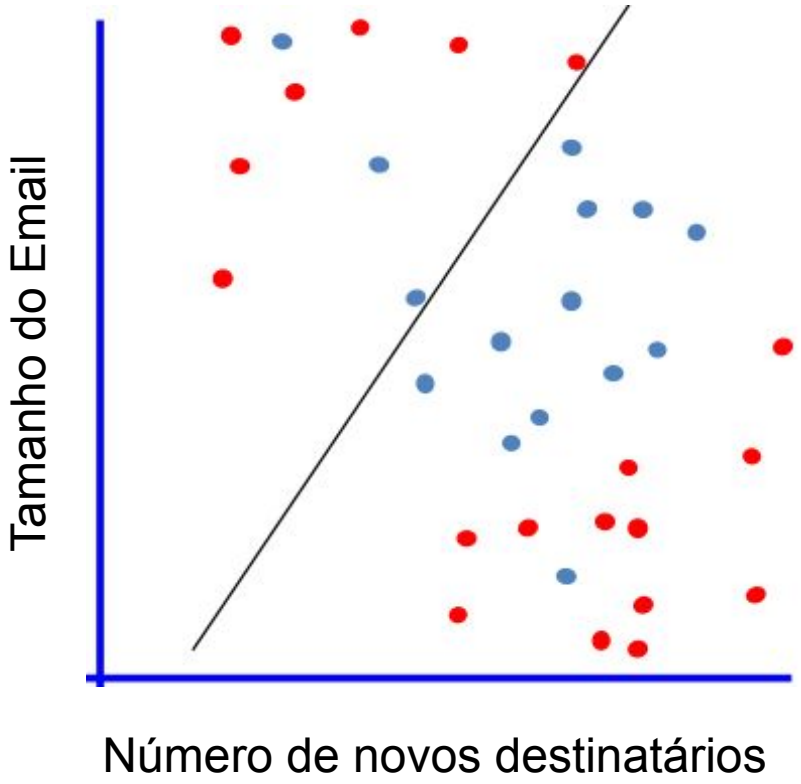
Matriz de Confusão

Real (Esperado) \ Classificado como	Azul	Verm.
Azul	7	1
Verm.	1	5

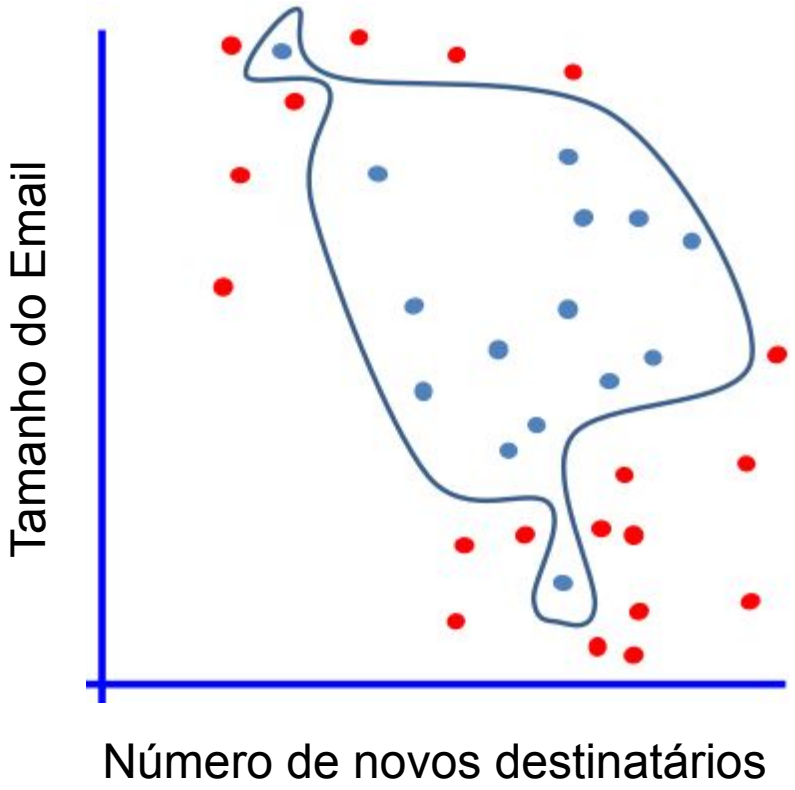
O Caso Não-linearmente Separável



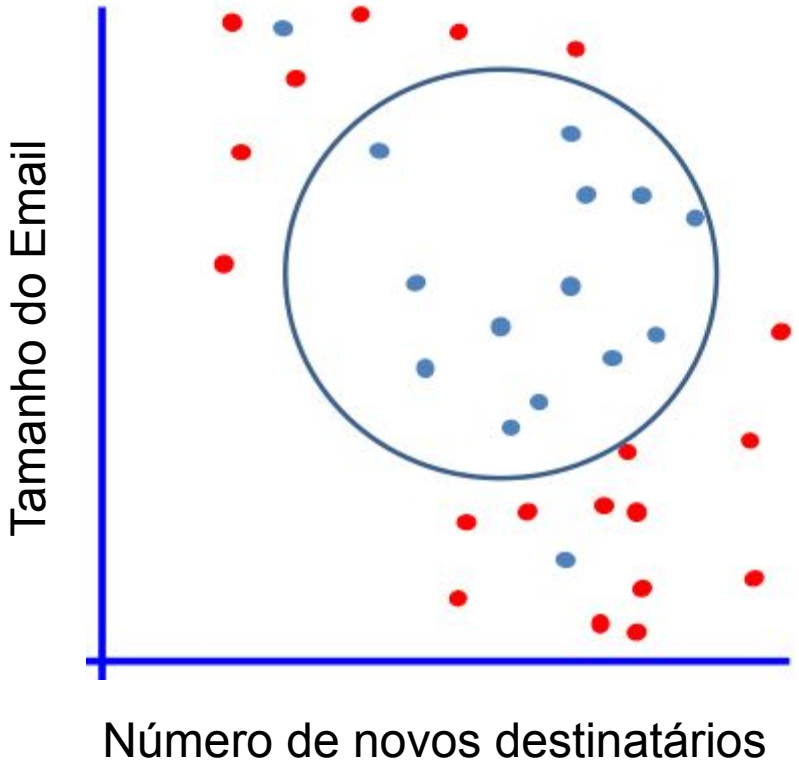
O Caso Não-linearmente Separável



O Caso Não-linearmente Separável

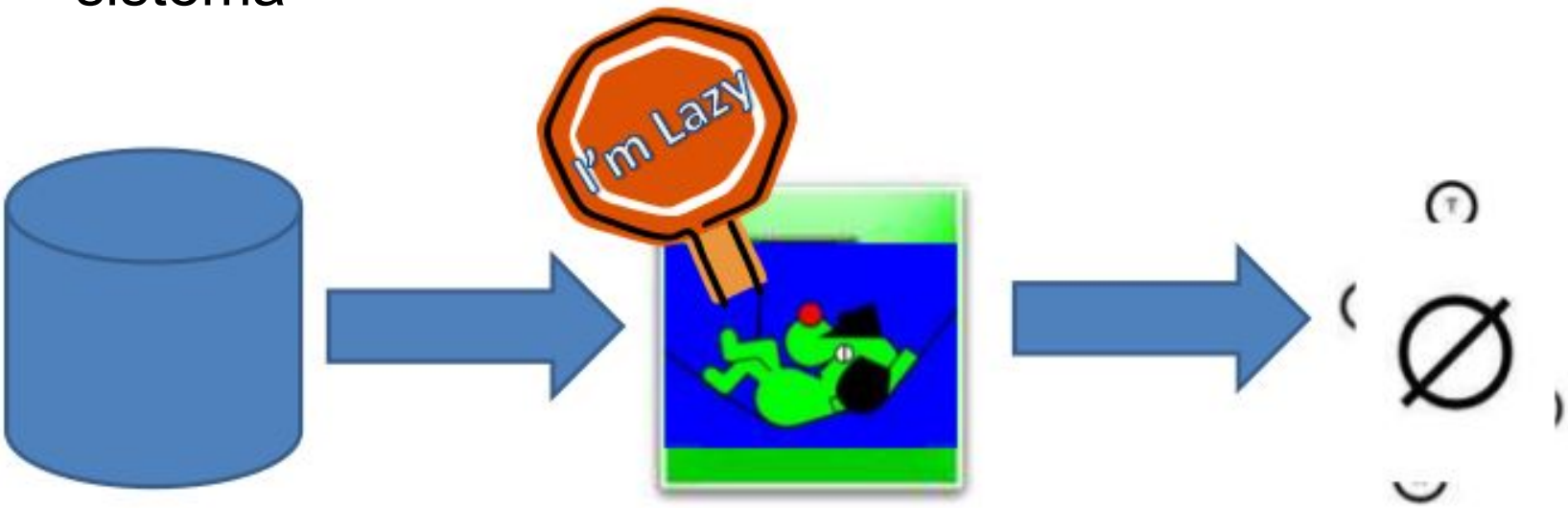


O Caso Não-linearmente Separável



Aprendizado *Lazy* (Preguiçoso)

- A generalização além do dados de treinamento é postergada até que uma nova instância é fornecida ao sistema



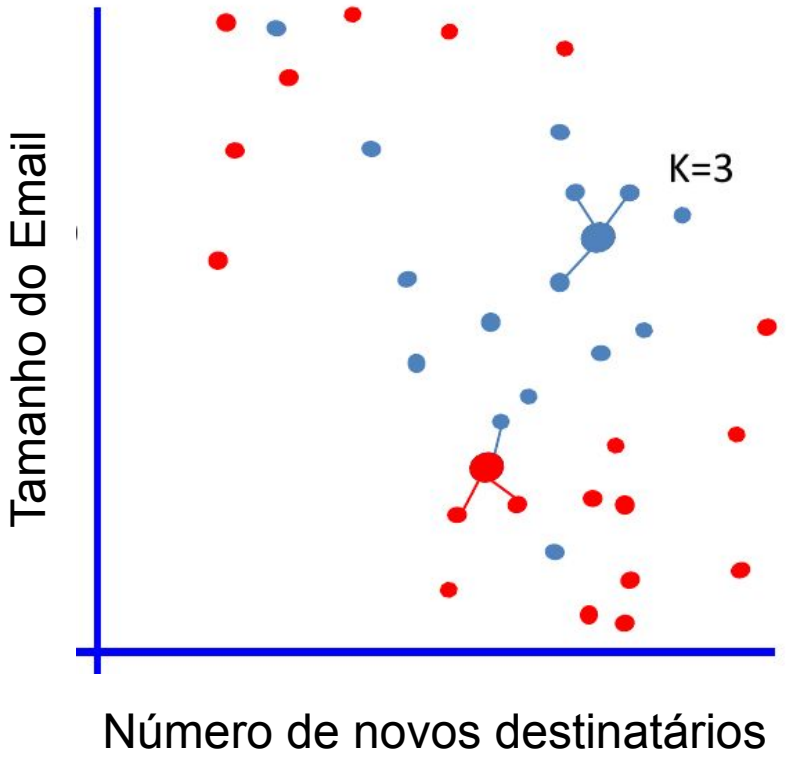
Aprendizado *Lazy*

Instance-based learning

Conjunto de
Treinamento

Aprendizado *Lazy*

K-Nearest Neighbors

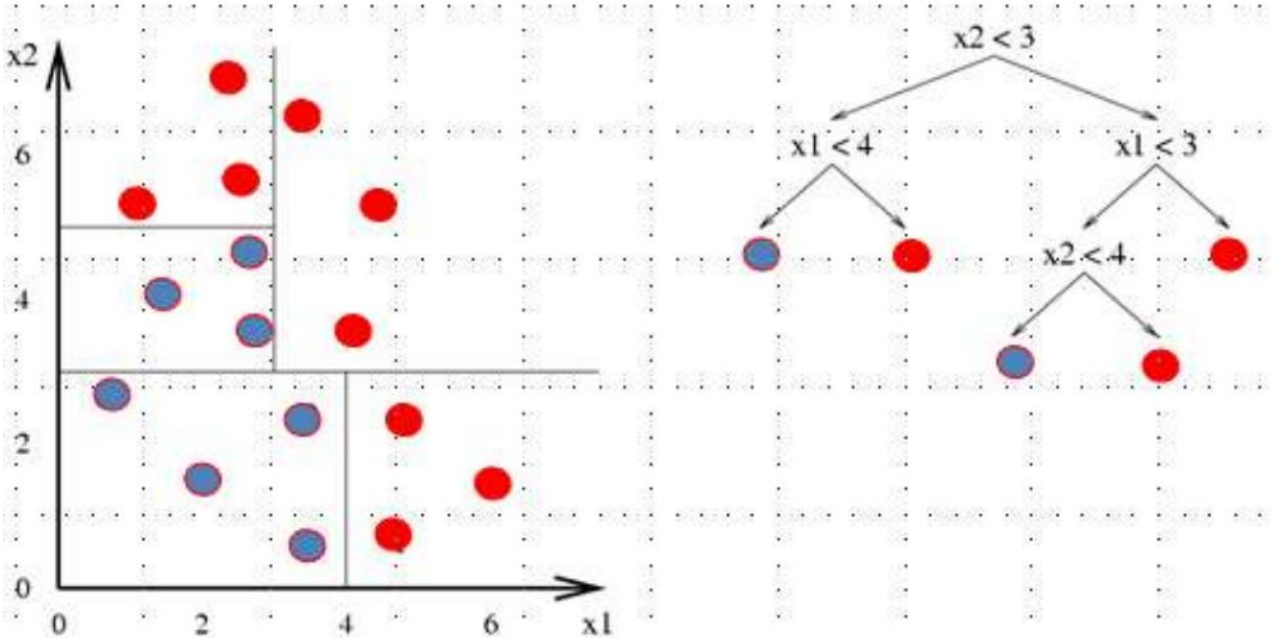


- Qual deve ser o K?
- Qual medida de distância deve ser usada?

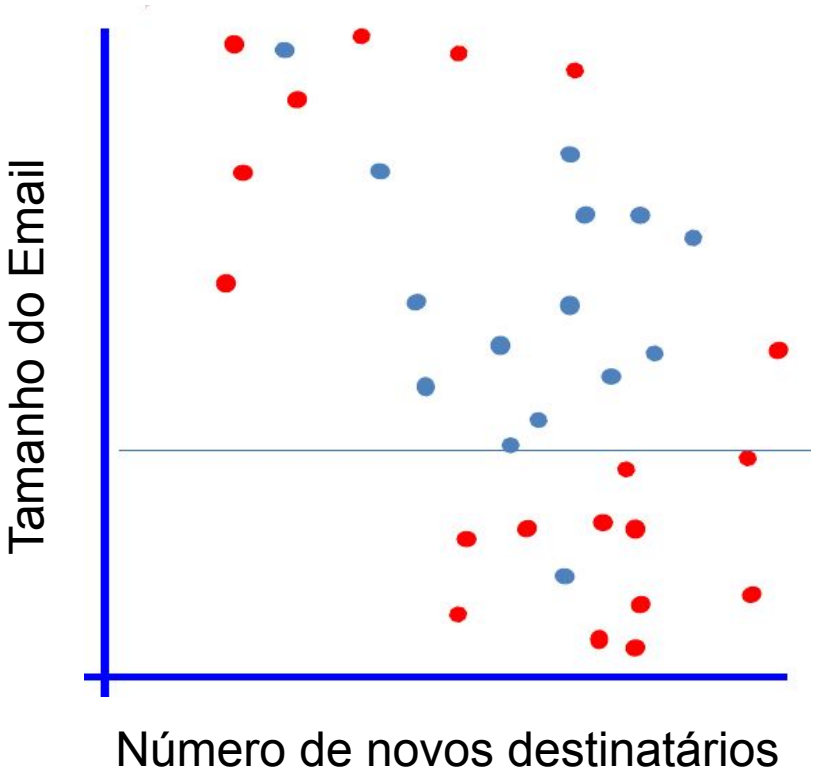
Árvores de Decisão

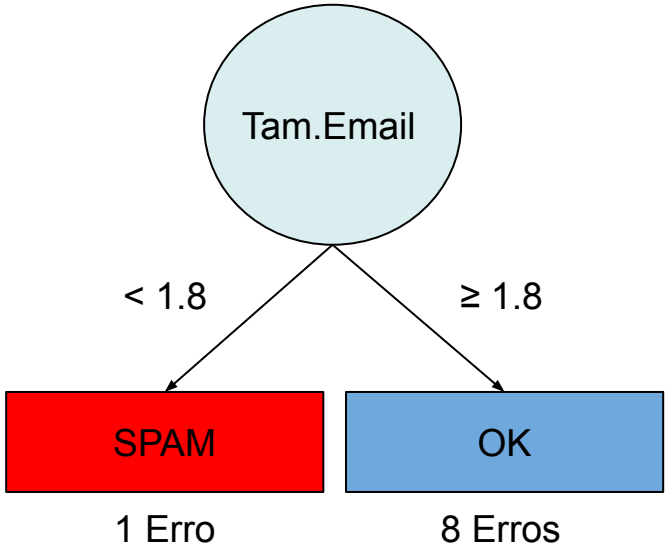
- Uma estrutura de árvore do tipo fluxograma
- Nós internos denotam uma avaliação em UM atributo
- Cada galho representa um resultado da avaliação
- Nós-folha representam uma classe/rótulo/meta

Árvores de decisão dividem o espaço de características em
eixos paralelos retangulares e rotulam cada retângulo com uma classe



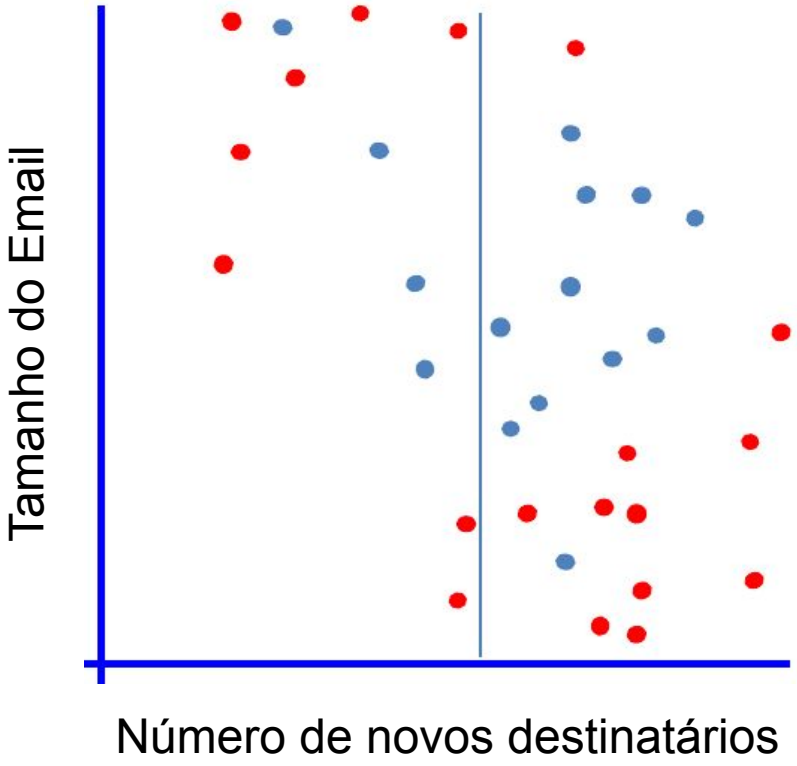
Indução *Top Down* de Árvores de Decisão

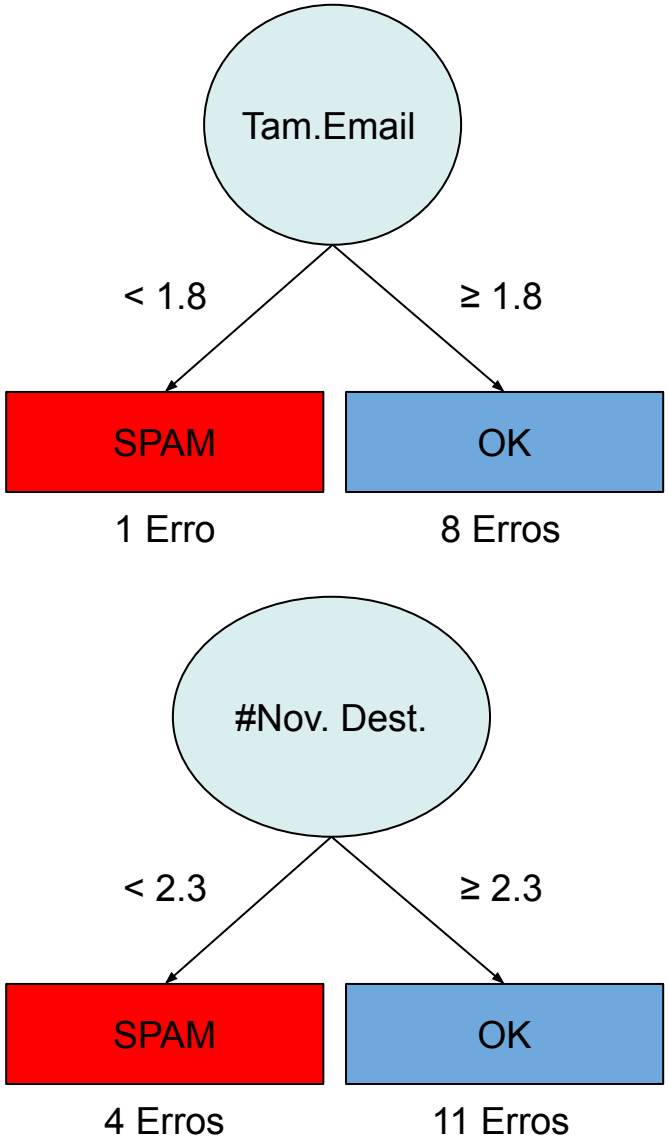




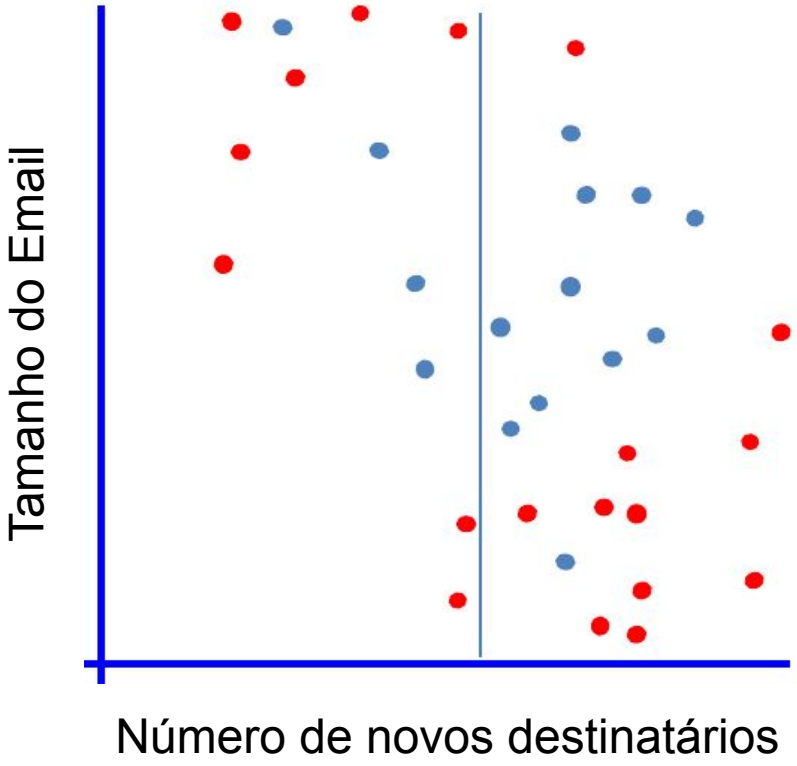
Uma árvore de decisão de um único nível é também conhecida como um Cepo/Toco-Decisão (*decision stump*)

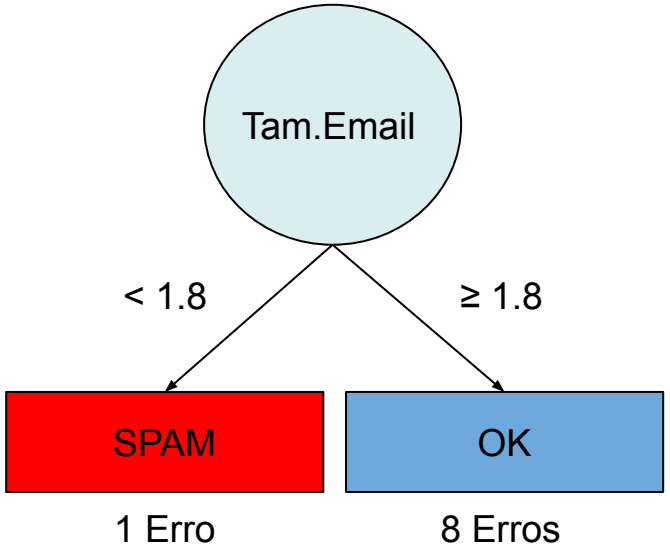
Indução *Top Down* de Árvores de Decisão

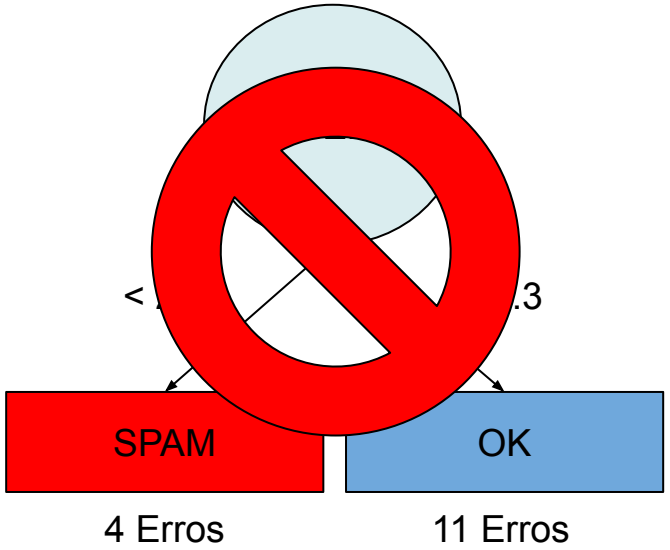




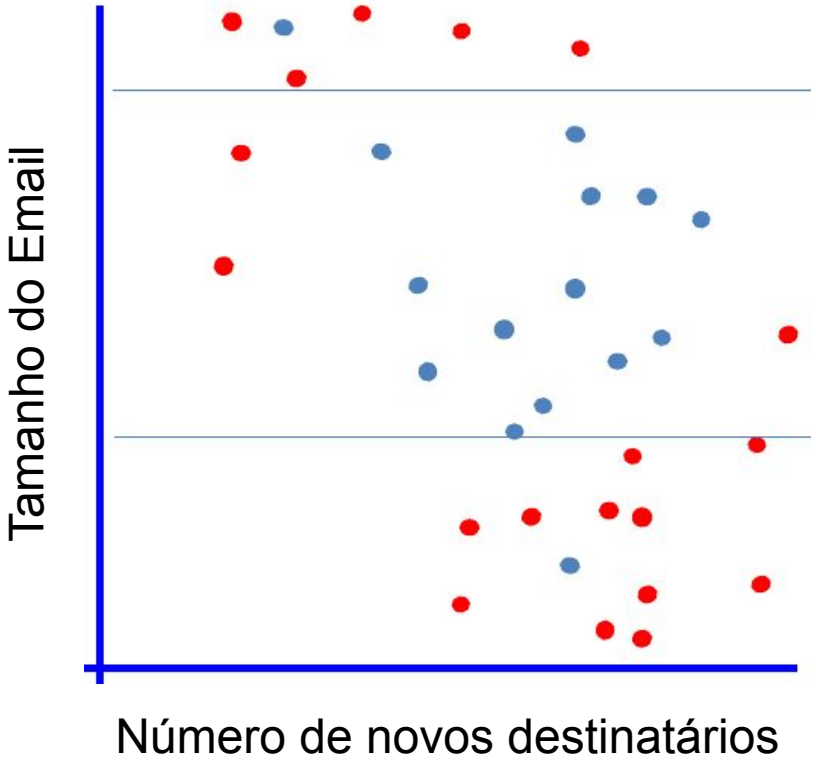
Indução *Top Down* de Árvores de Decisão

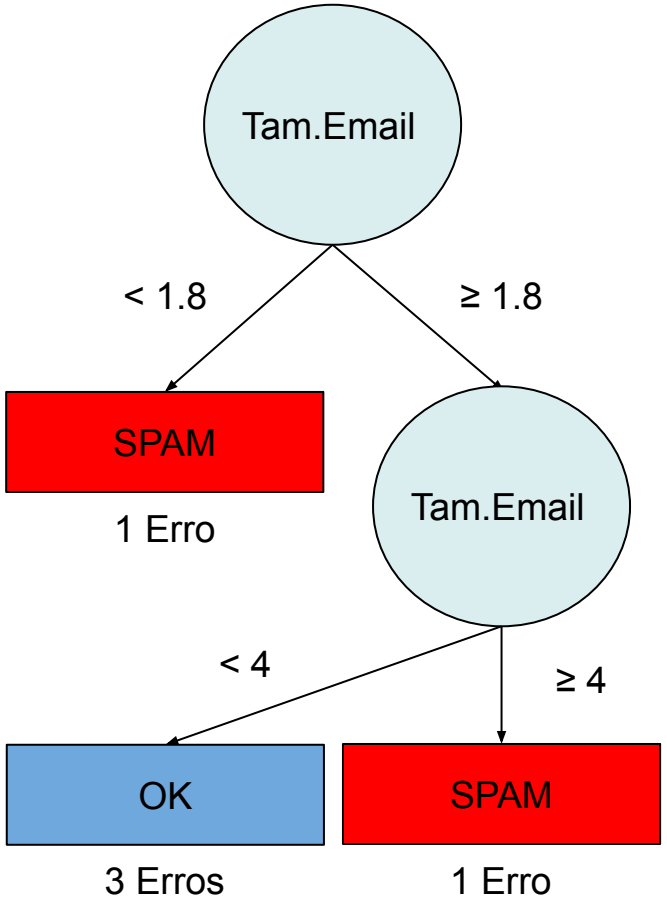




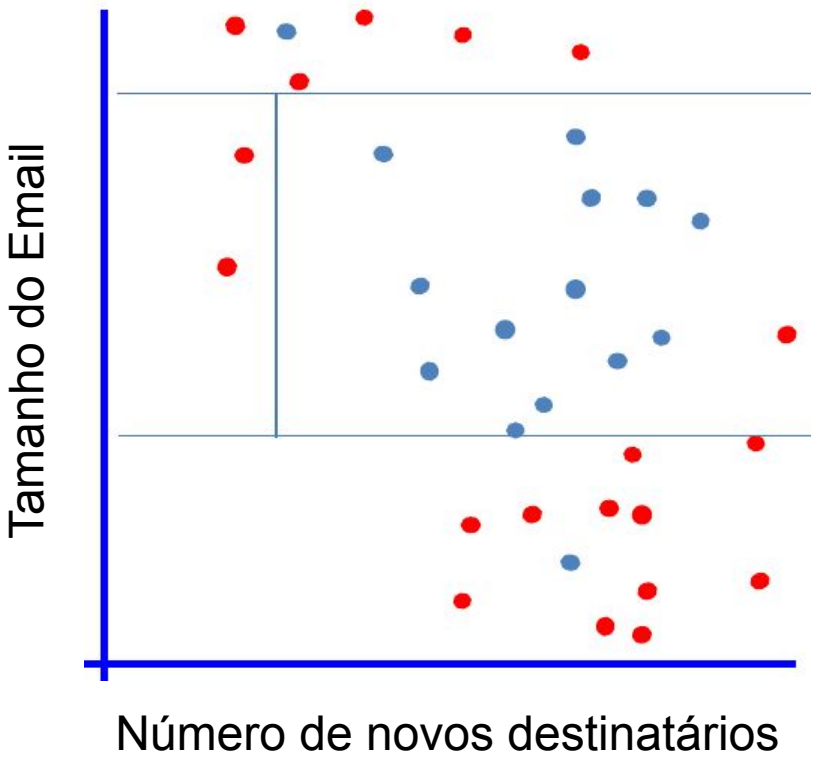


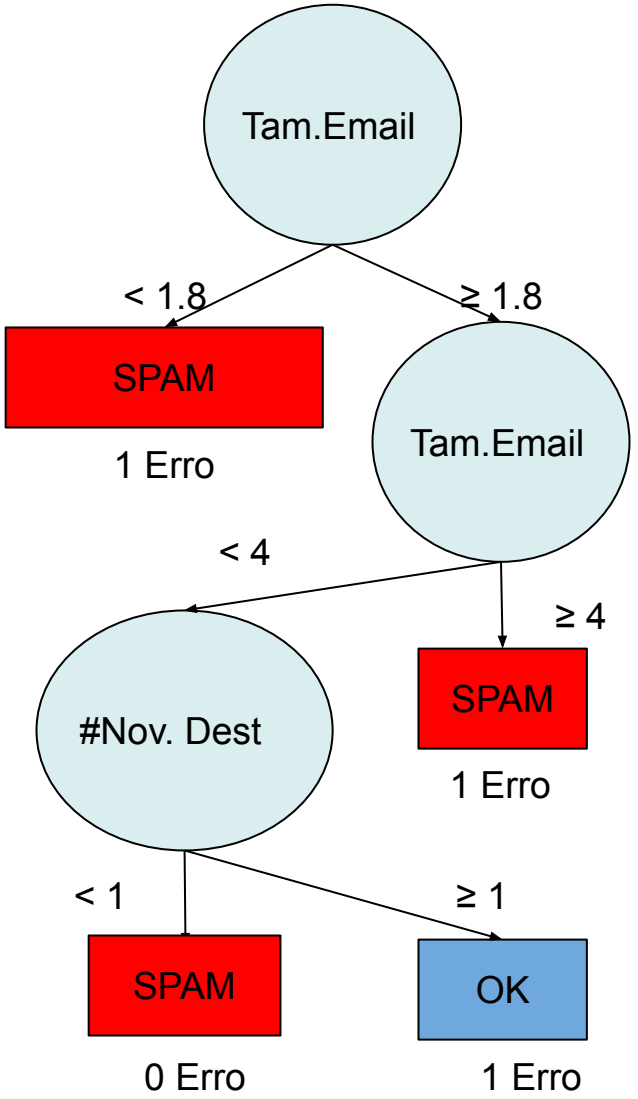
Indução *Top Down* de Árvores de Decisão



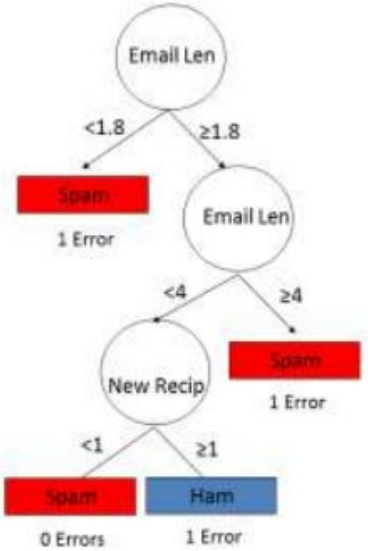


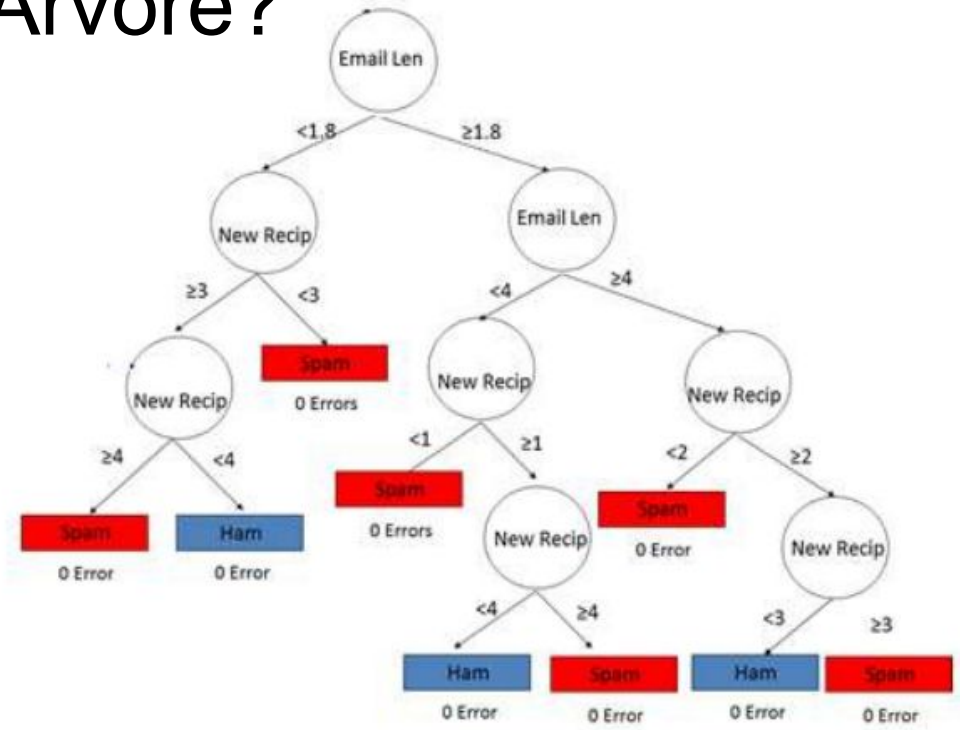
Indução *Top Down* de Árvores de Decisão

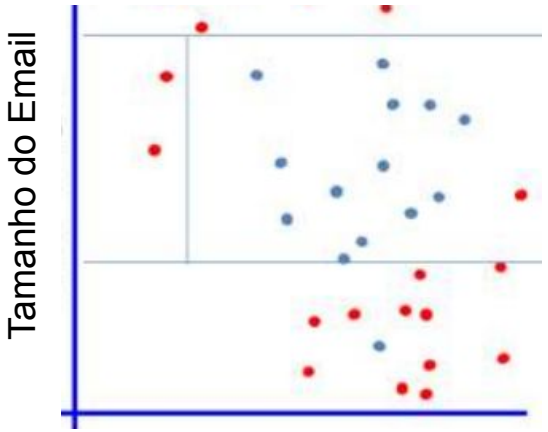


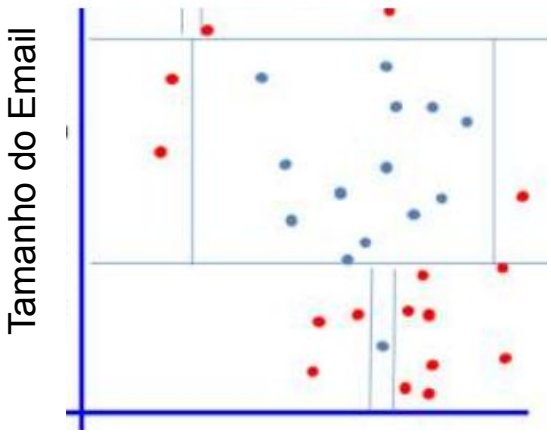


Qual Árvore?

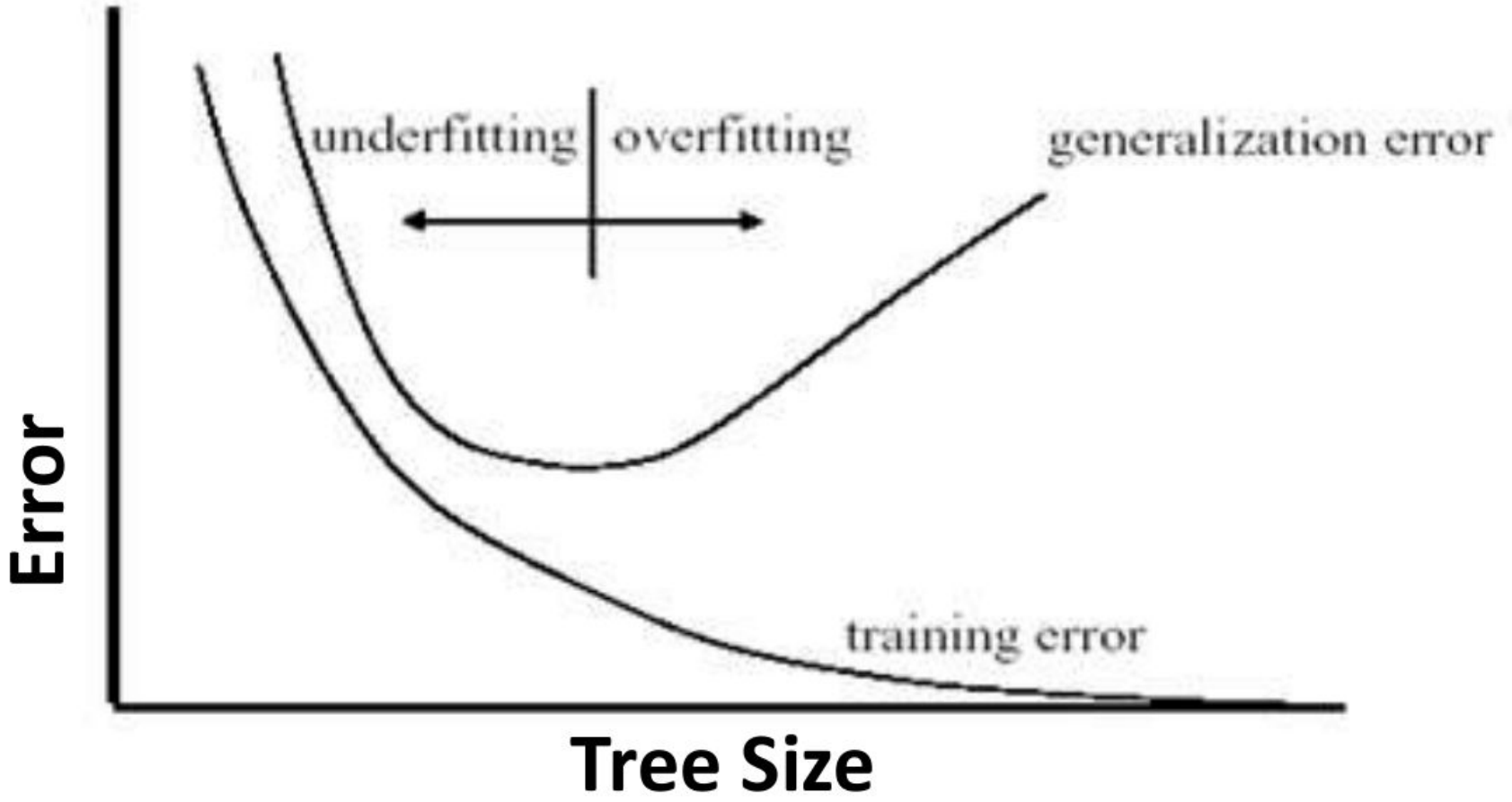

Email Len
<1.8
≥1.8
Spam
1 Error
Email Len
<4
≥4
New Recip
Spam
1 Error
<1
≥1
Spam
0 Errors
Ham
1 Error


Email Len
<1.8
≥1.8
New Recip
Email Len
≥3
<3
<4
≥4
New Recip
Spam
0 Errors
New Recip
New Recip
≥4
<4
<1
≥1
<2
≥2
Spam
0 Error
Spam
0 Errors
Ham
0 Error
New Recip
Spam
0 Error
New Recip
<4
≥4
<3
≥3
Ham
0 Error
Spam
0 Error
Ham
0 Error
Spam
0 Error


Tamanho do Email
de novos destinatários


Tamanho do Email
de novos destinatários

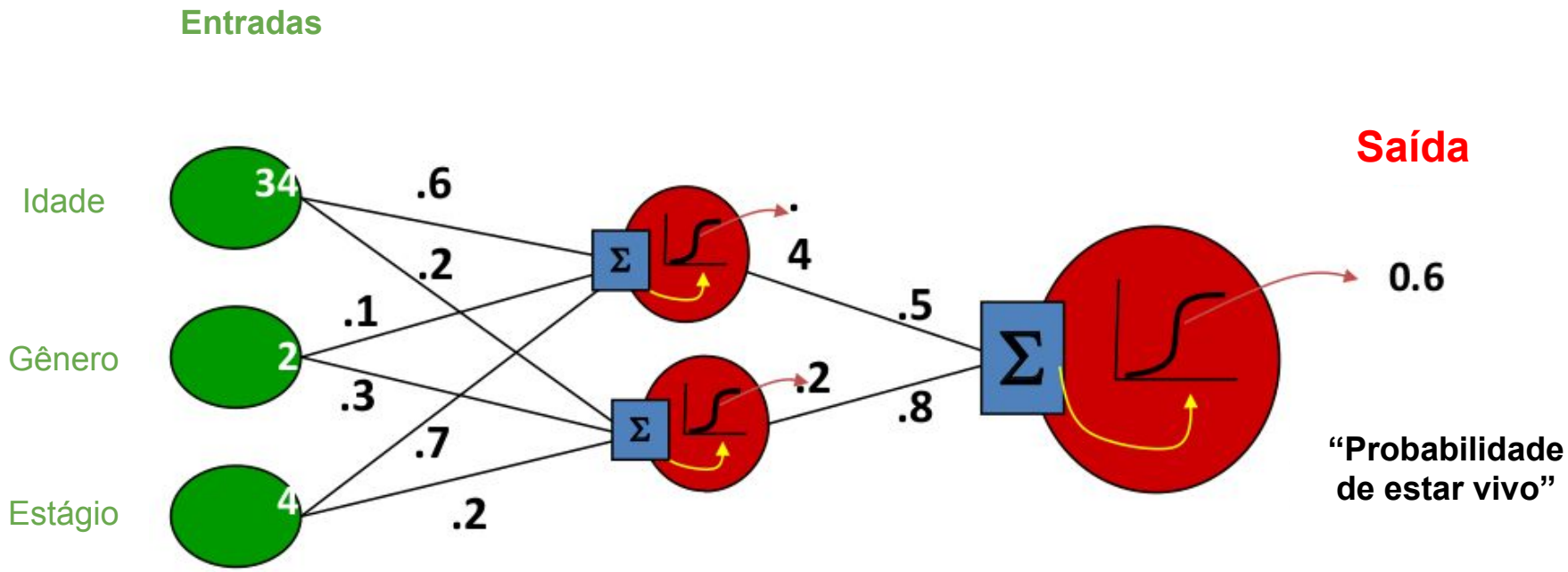
Overfitting & Underfitting



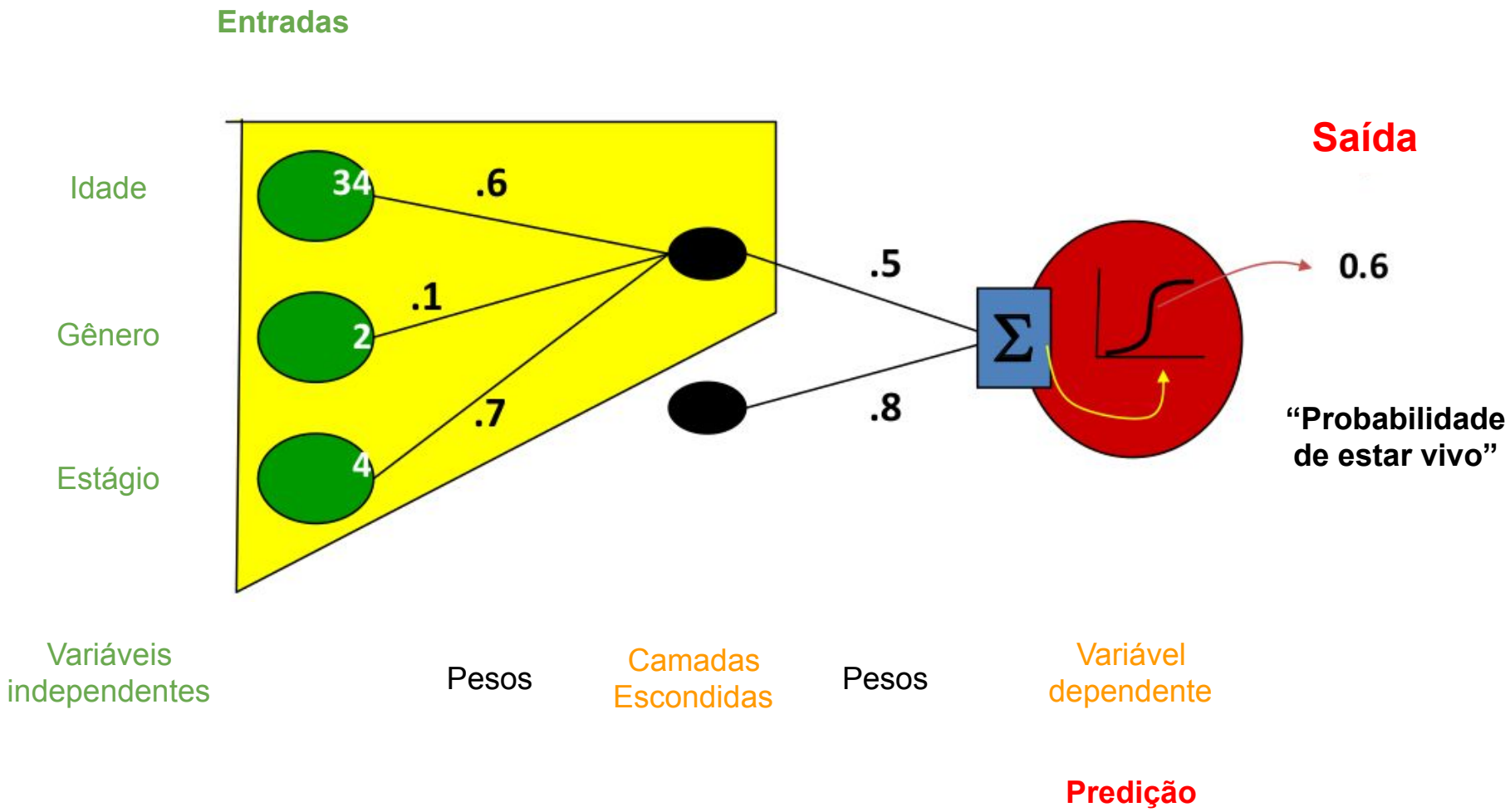
Overtraining: significa que ele aprendeu demais o conj. de treinamento (decorou) - ele se *superajustou* ao conj. de treinamento de forma que ele se desempenha mal no conj. de teste

Underfitting: significa que o modelo é muito simples, o erro no treino e no teste é muito grande

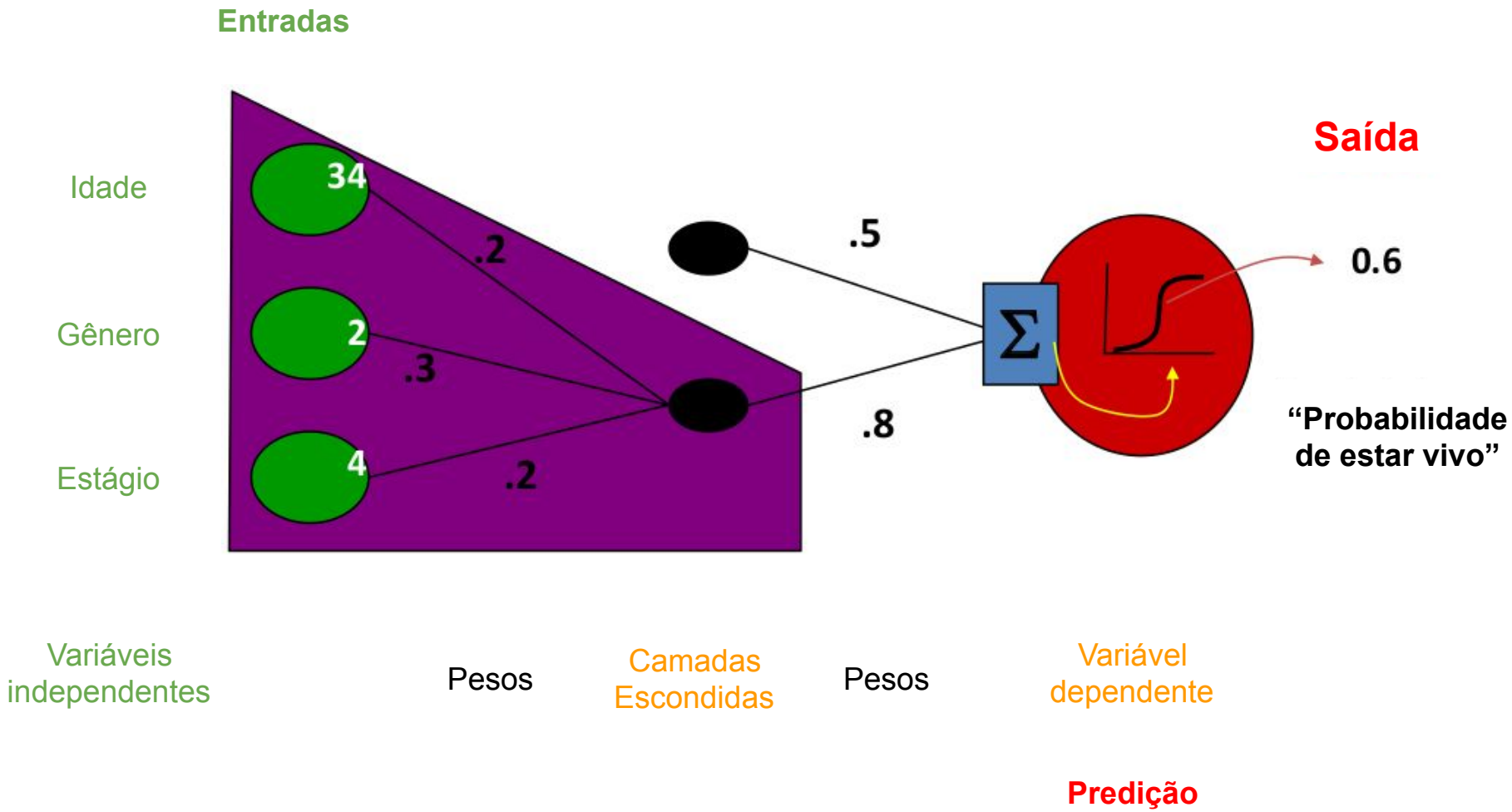
Modelo de Rede Neuronal



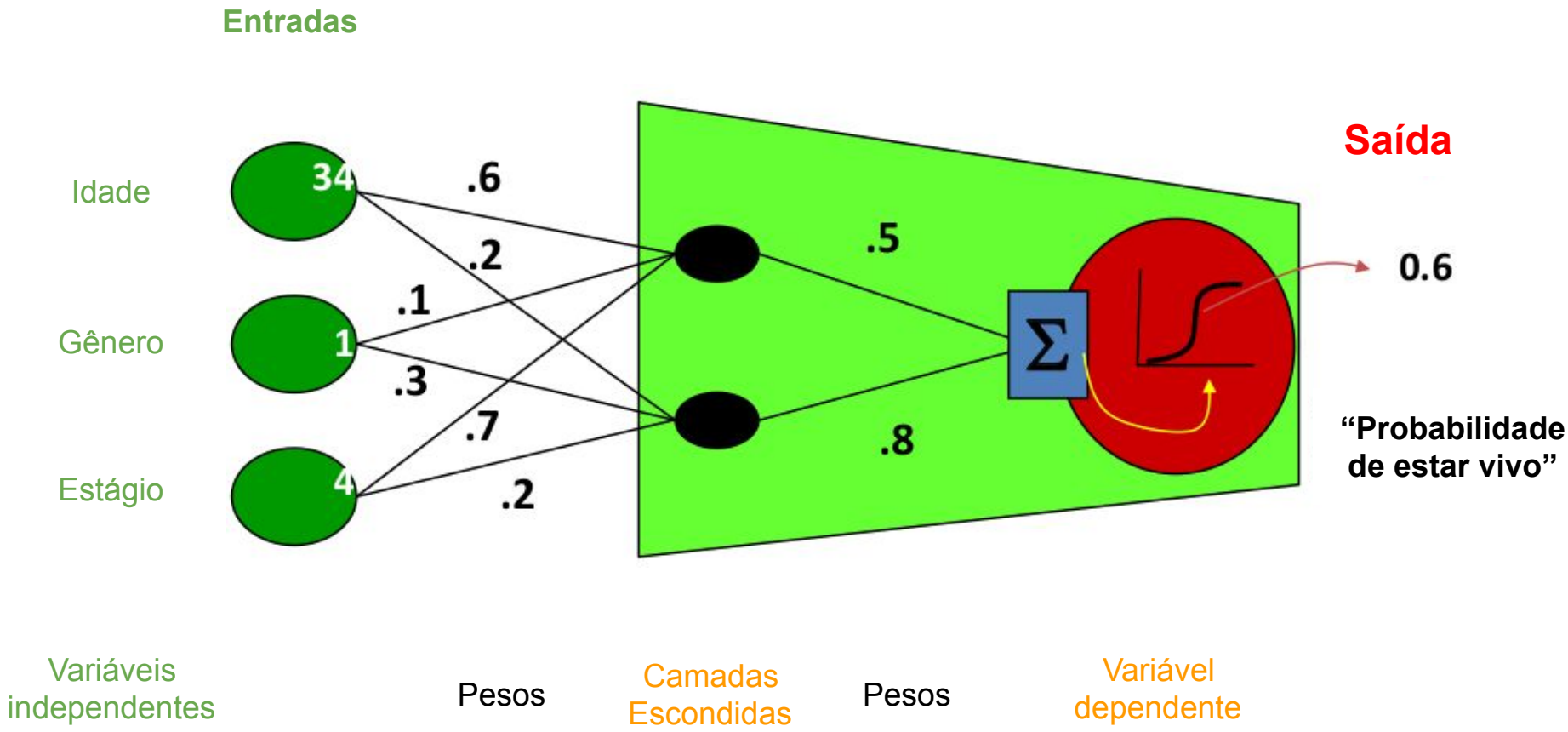
“Modelos Logísticos Combinados”



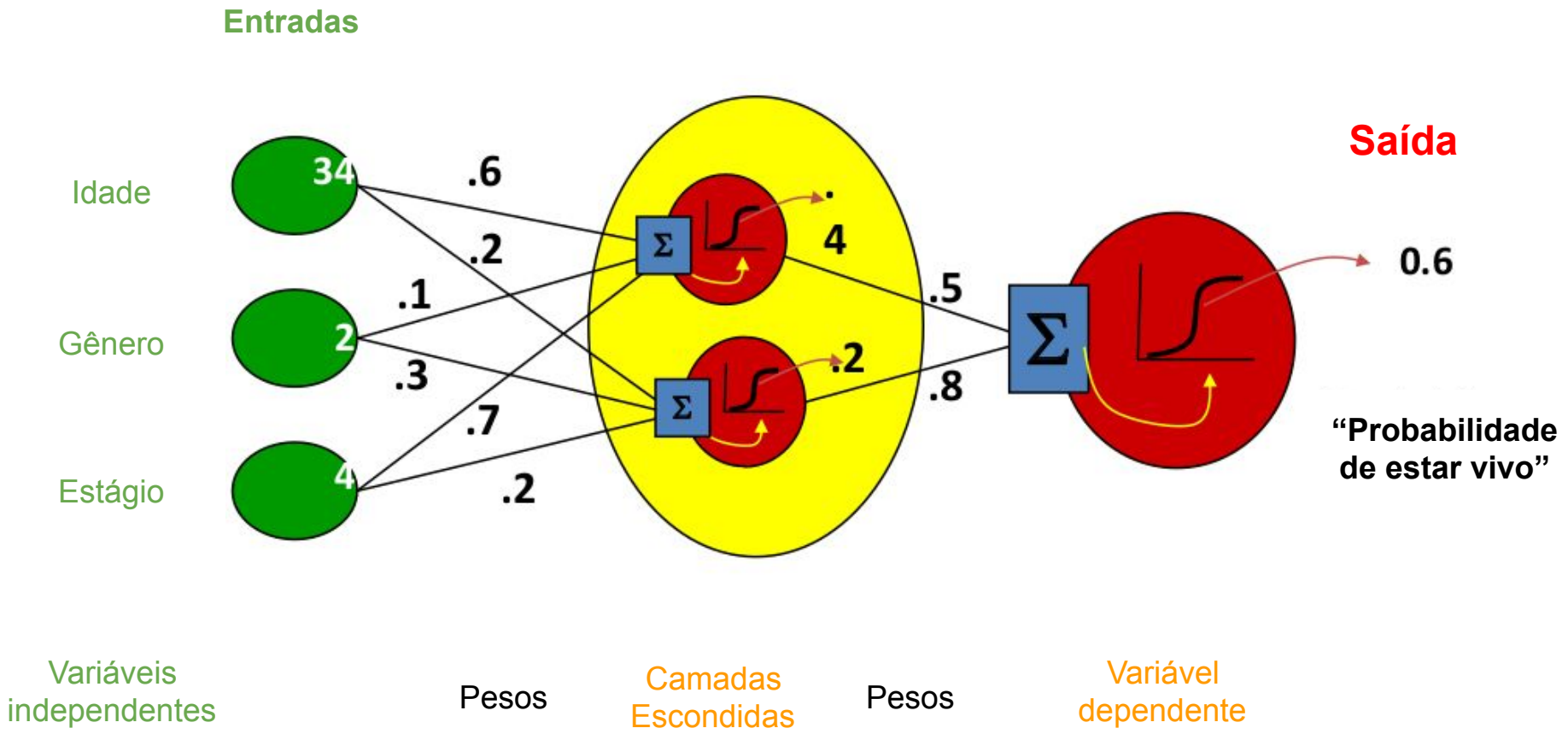
“Modelos Logísticos Combinados”



“Modelos Logísticos Combinados”



“Modelos Logísticos Combinados”



Aprendizado de *Ensembles*

- A ideia é usar múltiplos modelos para obter melhor desempenho preditivo que se poderia obter a partir de qualquer um dos modelos constituintes

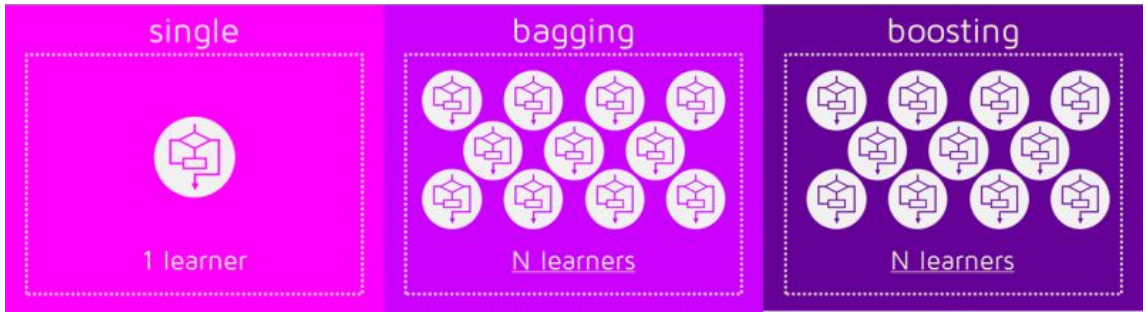
- Teoria da diversidade

Comitês
Combinação
Fusão

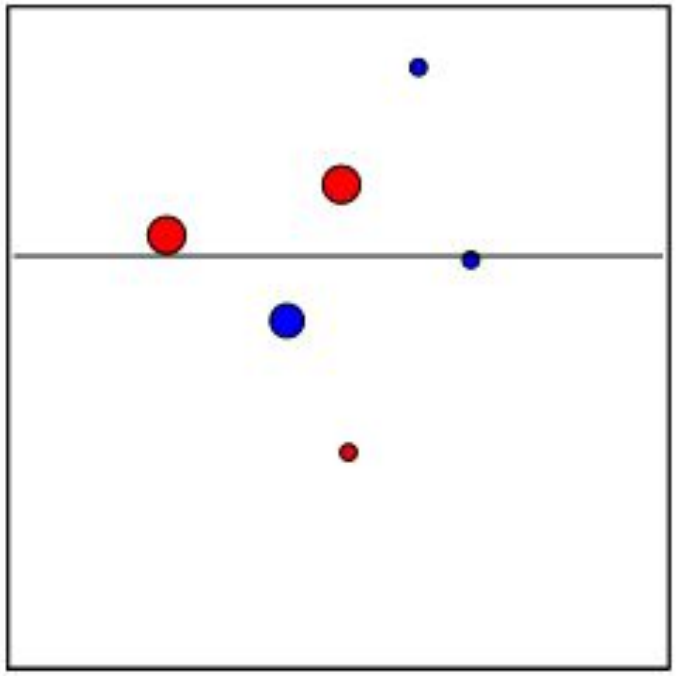
Aprendizado de *Ensembles*

- ***Bagging*** (*variance*) envolve a construção de um ***ensemble*** (comitê) treinando cada nova instância de modelo de classificação com conjuntos de treinamento diferentes (aleatoriamente escolhidos).

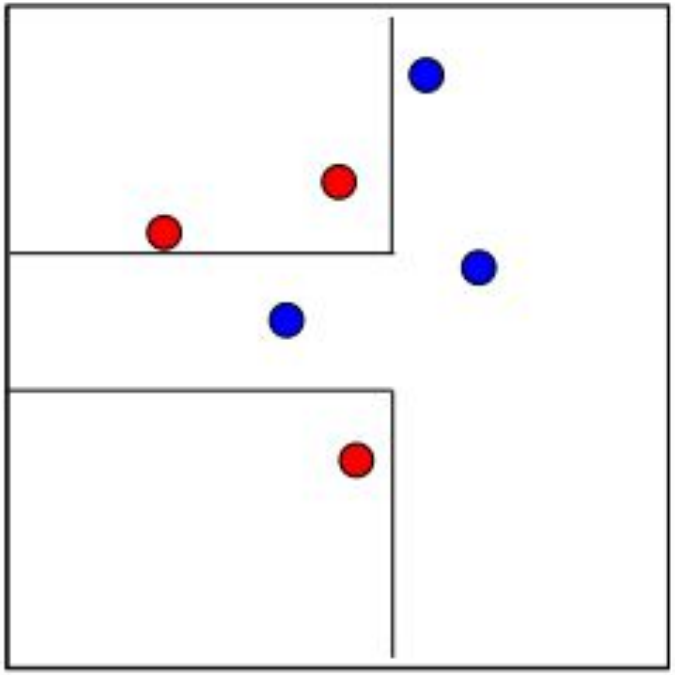
- ***Boosting*** (*bias*) ... para enfatizar as instâncias de treinamento que foram mal classificadas por modelos anteriores.



Exemplo de *Ensemble of Weak Classifiers*



Training



Combined classifier

Princípios Principais

Navalha de Occam (Século XIV)

- Do latim “ lex parsimoniae”
  - A lei da **parcimônia** (economia)
- A explicação de qualquer fenômeno deve fazer o menor número possível de suposições, eliminando aquelas que não fazem diferença nas predições observáveis da hipótese explicativa ou da teoria
- **O Dilema de Occam**: Infelizmente, em AM, acurácia e simplicidade estão em conflito.

No Free Lunch Theorem em Aprendizado de Máquina

- “ Para quaisquer dois algoritmos de aprendizado, há tantas situações (apropriadamente ponderadas) nas quais o algoritmo 1 é superior ao algoritmo 2 e vice-versa, de acordo com qualquer uma das medidas de ‘superioridade’ “

Então por que desenvolver novos algoritmos?

- O “desenvolvedor” (você) está preocupado em escolher o algoritmo mais apropriado para o **problema em questão**

- Isso requer algum conhecimento *a priori* - distribuição de dados, probabilidades anteriores, complexidade do problema, a física do fenômeno subjacente, etc.

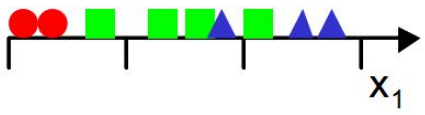
Então por que desenvolver novos algoritmos?

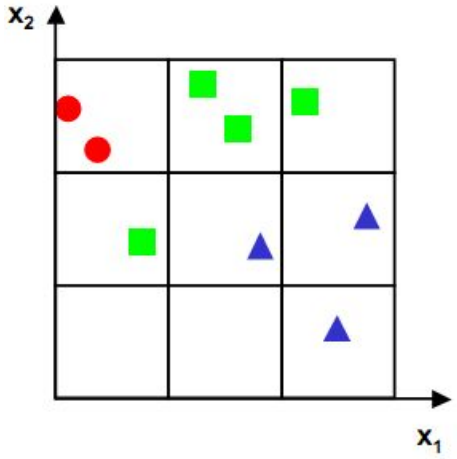
- O teorema do *No Free Lunch* nos diz que - a menos que tenhamos algum conhecimento *a priori* - classificadores simples (ou complexos) não são necessariamente melhores que outros. No entanto, dadas algumas informações *a priori*, certos classificadores podem melhorar as características de certos tipos de problemas.

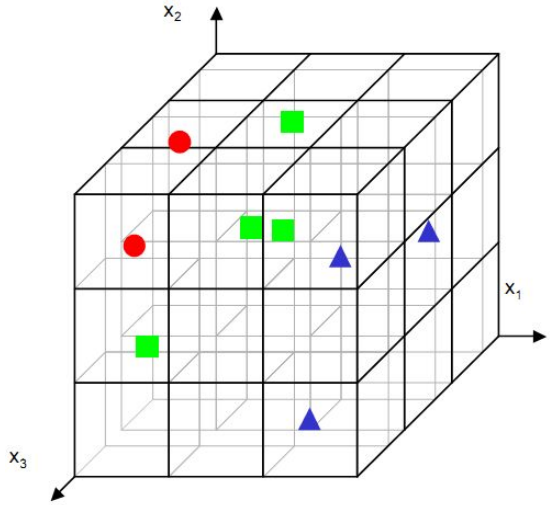
- O principal desafio do “desenvolvedor” é, então, identificar a correspondência correta entre o problema e o classificador! ... o que é mais um motivo para se armar com um conjunto diversificado / arsenal de algoritmos de aprendizado!

Menos é Mais

- A maldição da dimensionalidade (Bellman, 1961)







Menos é Mais
A maldição da dimensionalidade

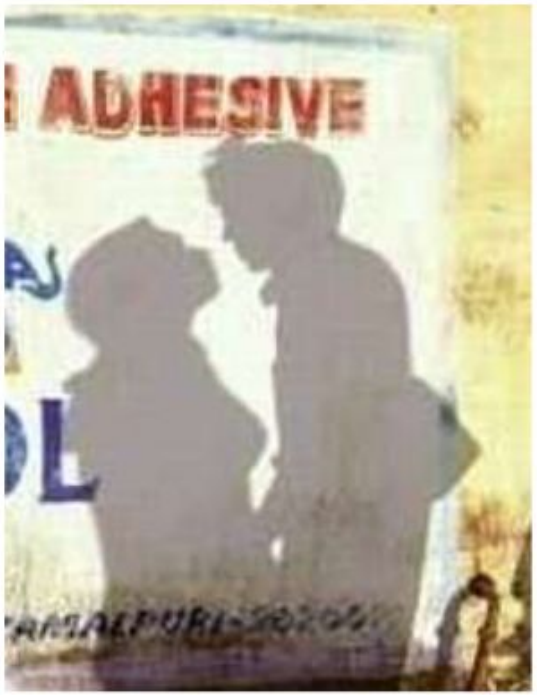
- Aprender a partir de um espaço de característica de alta dimensionalidade requer uma quantidade enorme de dados de treinamento para garantir que haja várias amostras com cada combinação de valores.

- Com uma quantidade fixa de número de instâncias de treinamento, o poder de preditibilidade reduz à medida que aumenta a dimensionalidade.

Menos é Mais
A maldição da dimensionalidade

- Como contra-medida, muitas técnicas de redução de dimensionalidade foram propostas, e foi demonstrado que, quando feitas adequadamente, as propriedades ou estruturas dos objetos podem ser preservadas.

- No entanto, aplicar ingenuamente a redução de dimensionalidade pode levar a resultados desastrosos.



- Como a **redução de dimensionalidade** é uma ferramenta importante no aprendizado de máquinas / mineração de dados / reconhecimento de padrões, deve-se estar sempre atento que ela pode distorcer os dados levando a representações enganosas.

- Acima temos uma projeção bidimensional de um mundo intrinsecamente tridimensional.

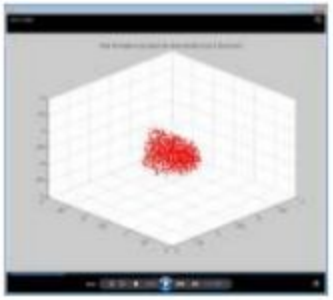
Fotógrafo desconhecido © Eammonn Keogh & Jessica Lin

- https://cs.gmu.edu/~jessica/DimReducDanger.htm

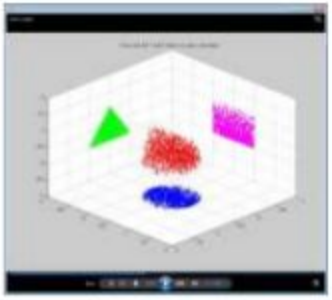
Redução da Dimensionalidade

- Recomenda-se usar o video original
  www.cs.gmu.edu/~jessica/DimReducDanger.htm

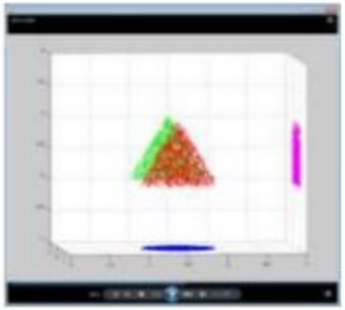
A cloud of points in 3D



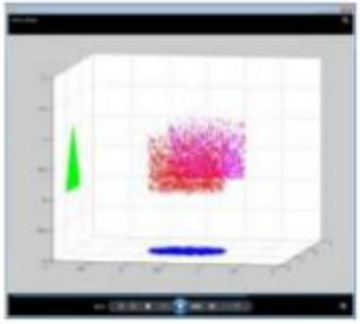
Can be projected into 2D
XY or XZ or YZ



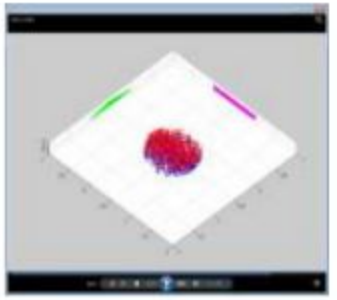
In 2D XZ we see
a triangle

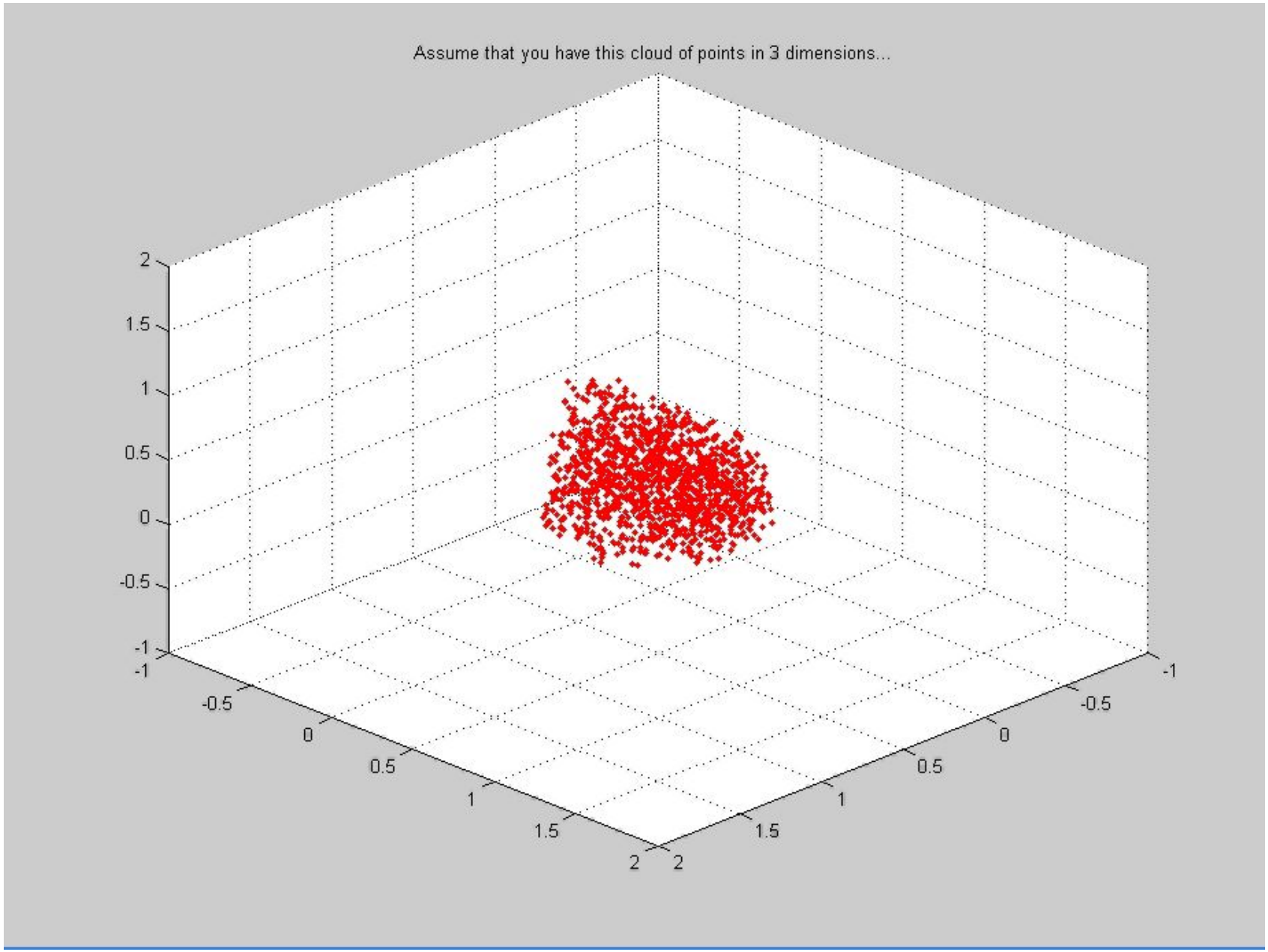


In 2D YZ we see
a square



In 2D XY we see
a circle





- www.cs.gmu.edu/~jessica/DimReducDanger.htm

Menos é Mais?

- No passado (2000-2010), o conselho disseminado era que a alta dimensionalidade era perigosa.
- Mas, reduzir dimensionalidade reduz a quantidade de informações disponíveis para previsão/aprendizado.
- Hoje: tenta-se ir na direção oposta: em vez de reduzir a dimensionalidade, aumenta-a adicionando muitas funções das variáveis preditoras.
- Quanto maior a dimensionalidade do conjunto de recursos, maior a probabilidade de ocorrer a separação (linear).

Significado das Respostas

- Um grande risco em mineração de dados é que você "descobrirá" padrões sem sentido.

- Os estatísticos chamam isso de princípio de Bonferroni: (grosseiramente) se você procurar em mais lugares por padrões interessantes do que sua quantidade de dados suportará, você está fadado a encontrar "porcaria".

Exemplos do Princípio de Bonferroni

- Rastreamento de Terroristas
- O paradóxo de Rhine: um grande exemplo de como não conduzir pesquisa científica

Por que o rastreamento de Terroristas é (Quase) Impossível?

- Suponha a crença de que certos grupos de malfeitores estão se encontrando ocasionalmente em hotéis para planejar a realização do mal.

- Deseja-se encontrar pessoas (não relacionadas) que tenham ficado pelo menos duas vezes no mesmo hotel no mesmo dia.

Os detalhes

- 10^9 (um bilhão) pessoas estão sendo rastreadas
- Por 1000 dias (≅3 anos)
- Cada pessoa fica em um hotel 1% do tempo (10 dias dos 1000)
- Hotéis acomodam 100 pessoas (então 10^5 hotéis)
- Se cada um comporta-se aleatoriamente (i.e., não malfeitores) a mineração de dados vai detectar alguma coisa suspeita?

As Contas (1)

- A probabilidade de que dado que as pessoas p & q vão estar no mesmo hotel em um dado dia *d*
  - $1/100 \times 1/100 \times 10^{-5} = 10^{-9}$.
- A probabilidade de p & q estejam no mesmo hotel nos dias d2 e d2:
  - $10^{-9} \times 10^{-9} = 10^{-18}$
- Pares de dias (combinação de 1000 tomados 2 a 2)
  - $\cong 5.10^{5}$.

As Contas (2)

- A probabilidade de p & q estarem no mesmo hotel em "algum" par de dias:
  - $5.10^{5} \times 10^{-18} = 5.10^{-13}$
- Pares de Pessoas:
  - 5.10^{17}
- Número esperado de pares de pessoas suspeitas:
  - $5.10^{17} \times 5.10^{-13} = 250\ 000.$

Conclusão

- Suponha que existem 10 pares de malfeitores que definitivamente estiveram no mesmo hotel duas vezes

- Os analistas/agentes tem que vascular entre 250.010 candidatos para encontrar os 10 casos reais
  - Nada vai acontecer
  - Como podemos melhorar este esquema?

Moral da História

- Quando estiver procurando por uma propriedade (e.g., “duas pessoas que estiveram em um mesmo hotel duas vezes”),

  certifique-se de que a propriedade não permite tantas possibilidades que dados aleatórios certamente produzam fatos "de interesse".

Paradóxo de Rhine (1)

- Joseph Rhine foi um parapsicólogo na década de 1950, que supunha que algumas pessoas tinham percepção extra-sensorial.

- Ele inventou (algo como) uma experiência em que os participantes foram solicitados a adivinhar 10 cartas escondidas - vermelho ou azul.

- Ele descobriu que quase 1 em 1000 tinha PES - eles conseguiram acertar todos os 10!

Paradóxo de Rhine (2)

- Ele disse a essas pessoas que elas tinham PES e as chamou para outro teste do mesmo tipo.

- Infelizmente, ele descobriu que quase todos eles haviam perdido sua PES.
  - Nenhuma pessoa acertou novamente?

- O que ele concluiu?
  - Responda no próximo slide.

Paradóxo de Rhine (3)

- Ele concluiu que você não deveria dizer às pessoas elas têm PES;
  - faz com que elas a percam.

Moral da História

- Entender o princípio de Bonferroni ajudará você a se parecer um pouco menos estúpido do que um Parapsicólogo

Instabilidade e o Efeito Rashomon

- Rashomon é um filme japonês em que quatro pessoas, de diferentes pontos de vista, testemunham um incidente criminal. Quando eles vêm para testemunhar no tribunal, todos relatam os mesmos fatos, mas suas histórias do que aconteceu são muito diferente.

- O efeito Rashomon é o efeito da subjetividade da percepção na lembrança (*recollection*).

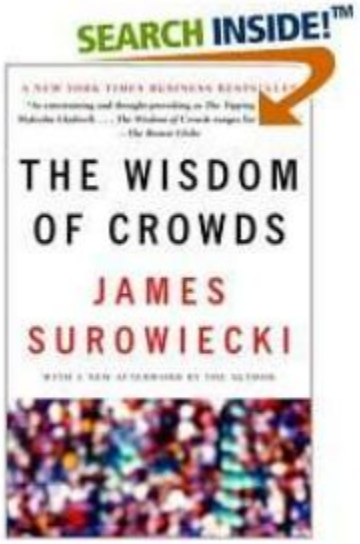
Instabilidade e o Efeito Rashomon

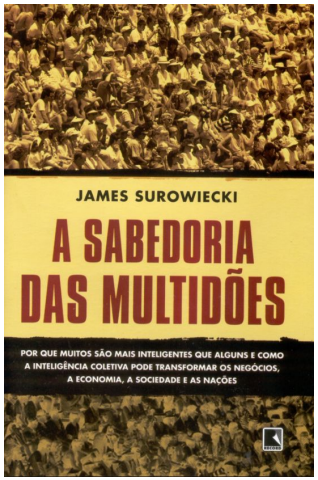
- O Efeito Rashomon em AM é que muitas vezes há uma multiplicidade de classificadores que dão a mesma taxa de erro mínima.

- Por exemplo, em árvores de decisão, se o conjunto de treinamento é perturbado apenas levemente, eu posso obter uma árvore bem diferente do original, mas com quase o mesmo erro do conjunto de testes.

A Sabedoria das Multidões

Por que Muitos são mais inteligentes que Alguns e como a Inteligência Coletiva pode transformar...





- Sob certas condições controladas, a agregação de informações em grupos, resultando em decisões que são frequentemente superior aos que podem ser feitos por qualquer um - até mesmo especialistas.

- Imita nossa segunda natureza de procurar várias opiniões antes de fazer qualquer decisão crucial. Nós pesamos as opiniões individuais e as combinamos para alcançar uma decisão final.

Comitês de Especialistas

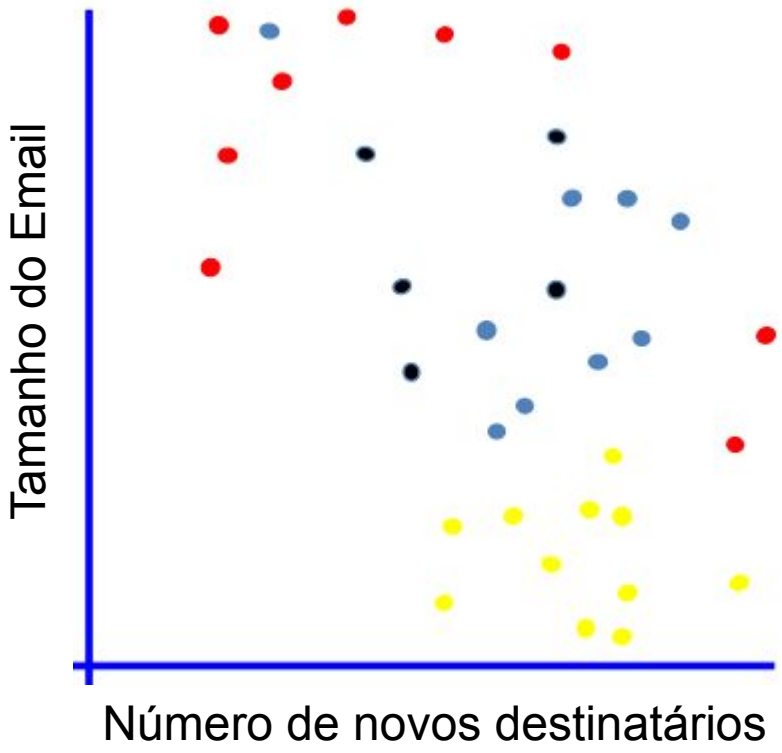
- "... uma escola de medicina que tem como objetivo que todos alunos, dado um problema, cheguem a uma solução idêntica"

- Não há muito sentido em criar um comitê de especialistas de tal grupo - tal comitê não melhorará o julgamento de um indivíduo.

- Considerar:
  - É preciso haver **desacordo** para o comitê ter o potencial de ser melhor que um indivíduo.

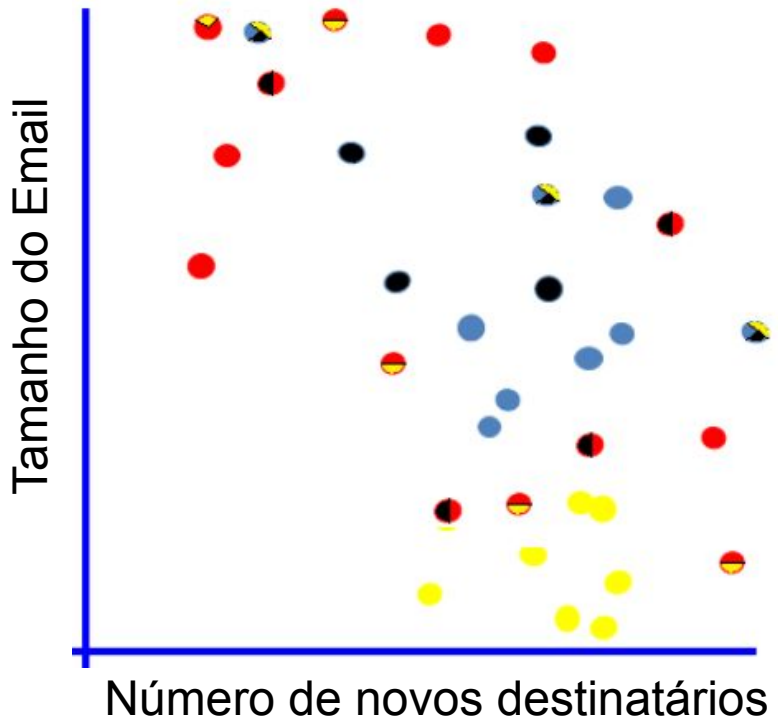
Outras Tarefas de Aprendizado

Aprendizado Supervisionado - *MultiClass*



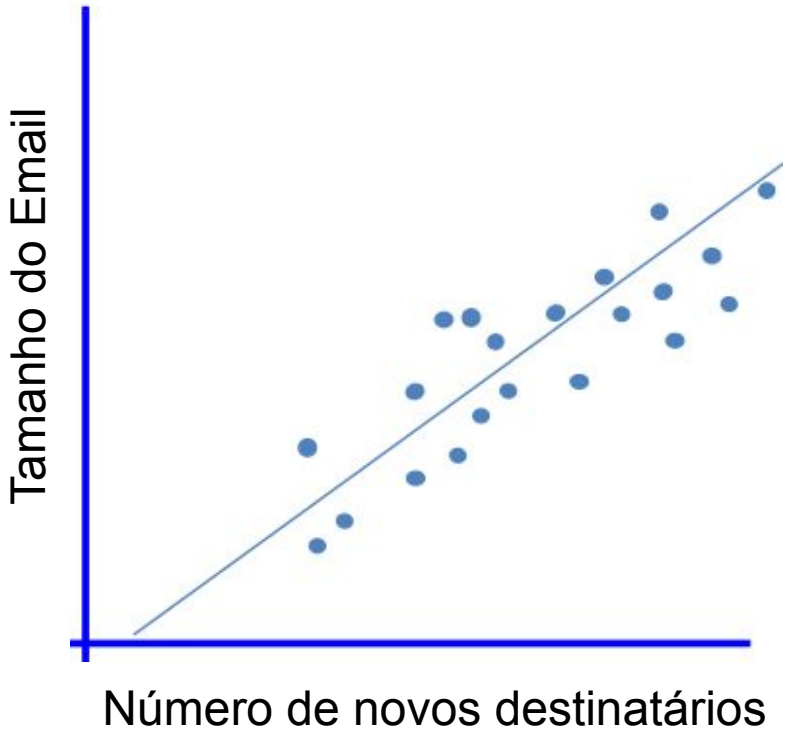
Aprendizado Supervisionado - *Multi-Label*

Multi-label learning refere-se ao problema de classificação onde cada exemplo pode ser atribuído a múltiplos rótulos simultaneamente



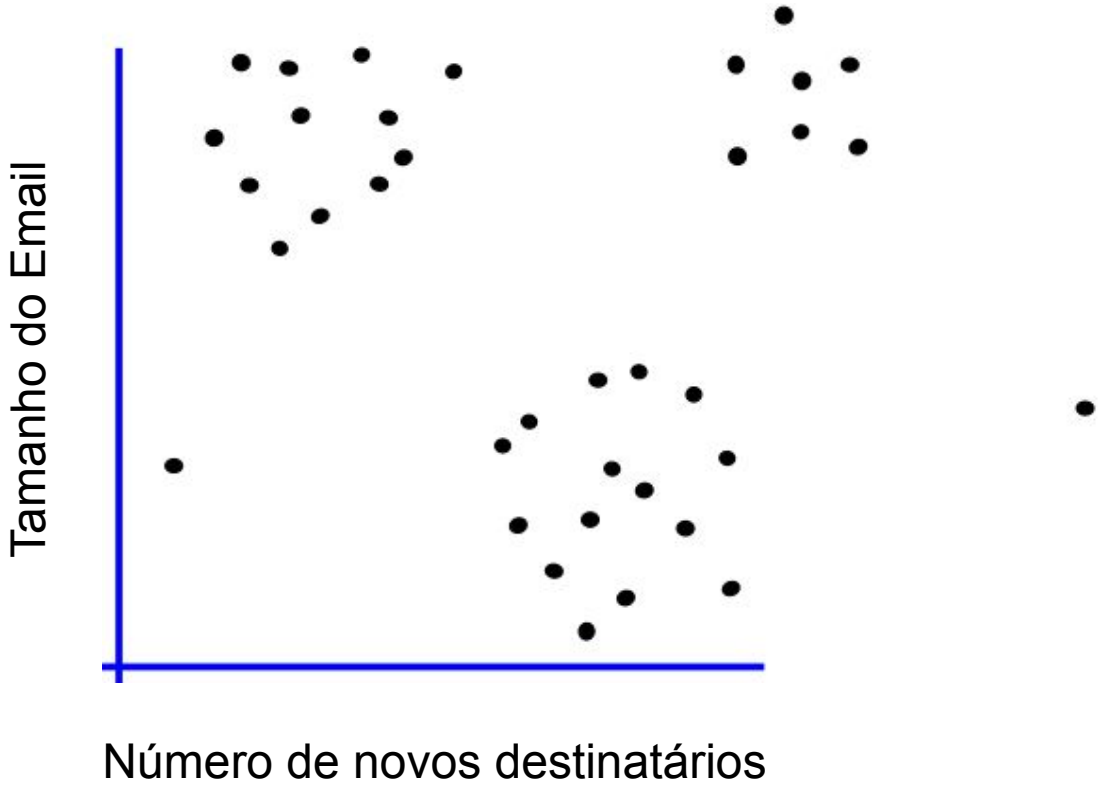
Aprendizado Supervisionado - **Regressão**

Encontrar uma relação entre uma variável dependente **numérica** e uma ou mais variáveis independentes



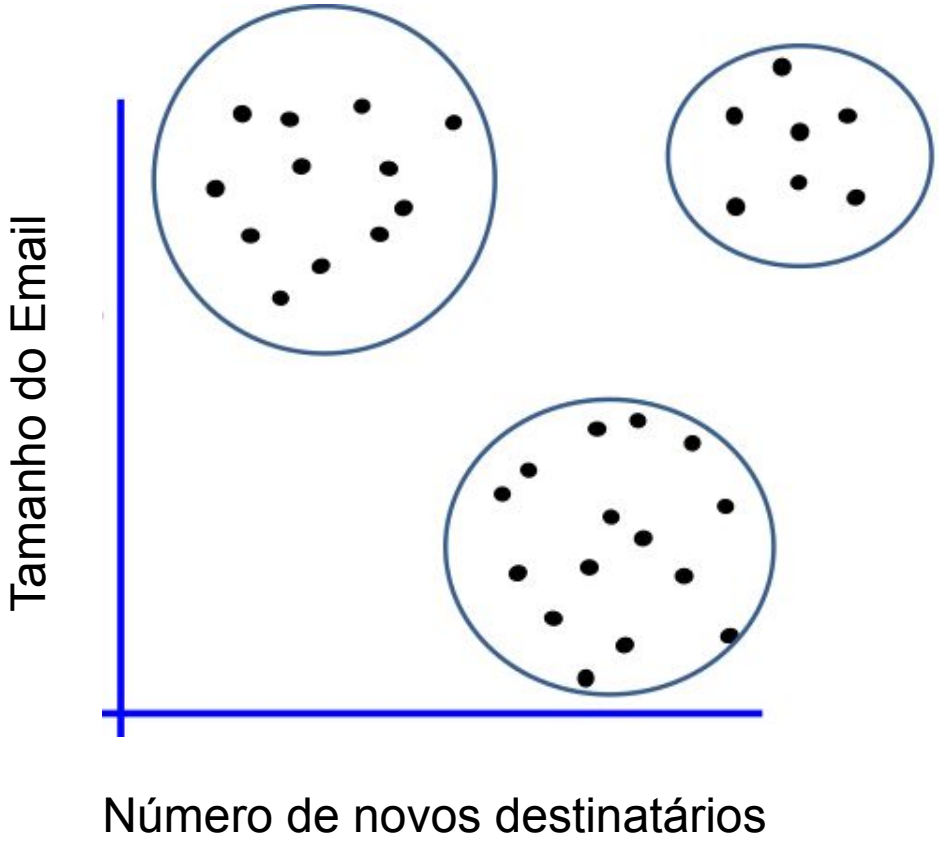
Aprendizado Não-Supervisionado
Detecção de Anomalias

Detectar padrões em um dado conjunto de dados que não se comportam / conformam a um comportamento padrão estabelecido.



Aprendizado Não-Supervisionado *Clustering*

Clustering é a designação de um conjunto de observações em subconjuntos (chamados *clusters* / aglomerados) de forma que as observações no mesmo cluster sejam similares (em algum sentido)



Fontes de Dados de Treinamento

- Fornecimento de exemplos aleatórios fora do controle do Algoritmo (*learner*).
  - Exemplos negativos disponíveis ou somente positivos?
    - Aprendizagem Semi-Supervisionada
  - Desbalanceamento de classes (atributo meta)

- O Algoritmo pode consultar um oráculo sobre a classe de um exemplo não rotulado no ambiente.
  - Aprendizado ativo

Fontes de Dados de Treinamento

- O Algoritmo pode construir um exemplo arbitrário e consultar um oráculo para seu rótulo.
  - Interpolação de exemplo vs Desembalancemento

- O Algoritmo pode executar diretamente no ambiente sem qualquer orientação humana e obter *feedback*.
  - Aprendizagem por reforço (*reinforcement learning*)

- Não há conceito de classe existente
  - Uma forma de descoberta
    - Aprendizagem não supervisionada
      - *Clustering*
      - Regras de Associação

Outras Tarefas de Aprendizado

- Outras configurações de aprendizado supervisionadas
  - *Multi-class Classification*
  - *Multi-label Classification*
  - *Semi-supervised Classification*
    - uso de dados rotulados e não rotulados
  - *One Class Classification* - somente instâncias de um rótulo são fornecidas

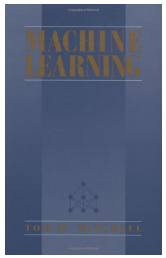
- Aprendizagem de *Ranking* e *Preferências*
  - Máquinas de Busca
  - Máquinas de Recomendação

Outras Tarefas de Aprendizado

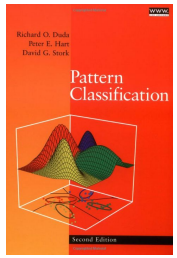
- Aprendizado *on-line* e Aprendizado incremental
  - Aprende uma instância por vez.
- *Concept drift*
- *Multi-task Learning & Transfer Learning*
  - *DeepLearning*
- Classificação coletiva - quando as instâncias são dependentes!
  - Multiple-instance learning

Software

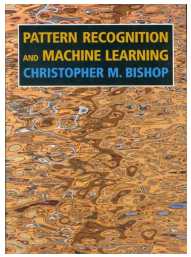
Quer Aprender Mais?



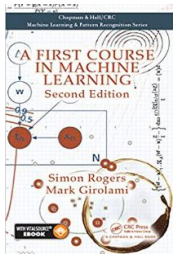
- T. Mitchell, **Machine Learning**, McGraw Hill, 1997.
  - http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html



- R. Duda, P Hard and D. Stork, **Pattern Classification**, Wiley-Interscience, 2000.



- C. M. Bishop, **Pattern Recognition and Machine Learning**, Springer, 2006.



- Simon Rogers and Mark Girolami. **A First Course in Machine Learning - second edition** Chapman & Hall/CRC, 2016.

Agenda

- Demais aulas
  - 05/10 - 2 aulas / 4 horas
  - 19/10 - 4 aulas / 8 horas
  - 09/11 - 4 aulas / 8 horas

Referências

- Notas de aulas do Prof. Lior Rokach
  - liorrk@bgu.ac.il - http://www.ise.bgu.ac.il/faculty/liorr/
- MOOC - Massive online-courses
  - https://www.class-central.com
  - edX
  - Stanford University (coursera) (Andrew Ng)
  - Udacity

Nota Importante

- Os slides desta aula são uma tradução com adaptações das notas de aulas do Prof. **Lior Rokach**
  - **Email:** liorrk@bgu.ac.il
  - **Homepage**: http://www.ise.bgu.ac.il/faculty/liorr/
  - **Slides**: https://pt.slideshare.net/liorrokach/introduction-to-machine-learning-13809045