

Data science aplicada aos negócios

Prof. Dr. Álvaro Campos
Ferreira

alvaro.ferreira@idp.edu.br

Aprendizado de Máquina

Instância vs Modelo



Imagine uma forma de escrever um programa para detectar e-mails que sejam Spam.

Como vocês resolveriam esse problema?

Aprendizado de Máquina

Já está entre nós, desde a década de 1990 e duas aplicações correntes são:

- Reconhecimento óptico de caracteres (OCR)
- Filtro de Spam

Aprendizado de Máquina



Aprendizado de Máquina é a arte e a ciência de programar computadores para que eles possam aprender com dados.

Aprendizado de Máquina

Aprendizado de Máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados.

Arthur Samuel, 1959.

Aprendizado de Máquina

É dito que um programa de computador aprende com a Experiência E com relação a alguma Tarefa T e uma medida de Performance P , se sua performance em T , como medido por P , melhora com a experiência E .

Tom Mitchell, 1997.

Aprendizado de Máquina

O filtro de Spam é um programa que utiliza exemplos de e-mails rotulados como Spam para marcar novos e-mails como Spam ou não.

O nome do conjunto de dados que o sistema usa para aprender é conjunto de dados de treinamento (training data)

Aprendizado de Máquina

O filtro de Spam é um programa de Aprendizado de Máquina que utiliza exemplos de e-mails rotulados como Spam para marcar e-mails como Spam:

- Tarefa T: Marcar Spam
- Experiência E: Dados de treinamento (e-mails)
- Performance P: Precisa ser definida

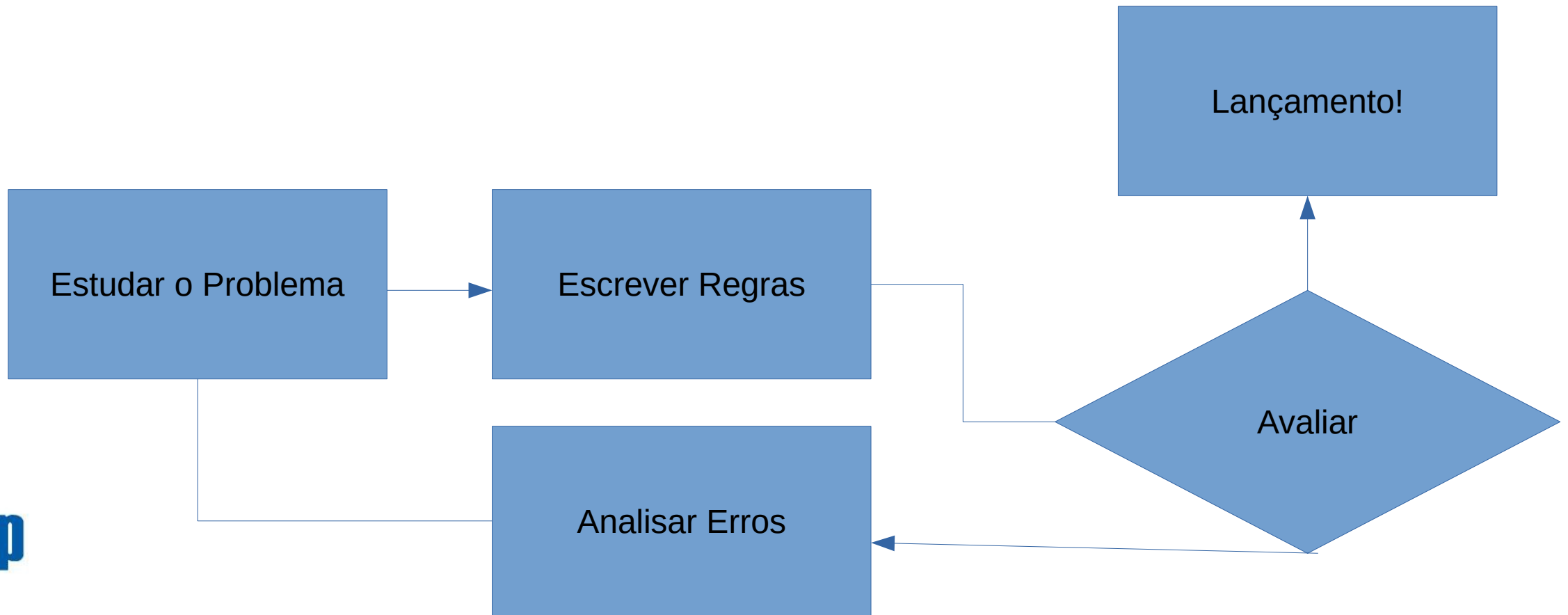
Aprendizado de Máquina

A Performance P do sistema de Aprendizado de Máquina deve ser definida para o caso estudado. Por exemplo, podemos usar a relação de e-mails classificados corretamente.

- Essa medida é chamada precisão

Por que utilizar Aprendizado de Máquina?

A abordagem tradicional

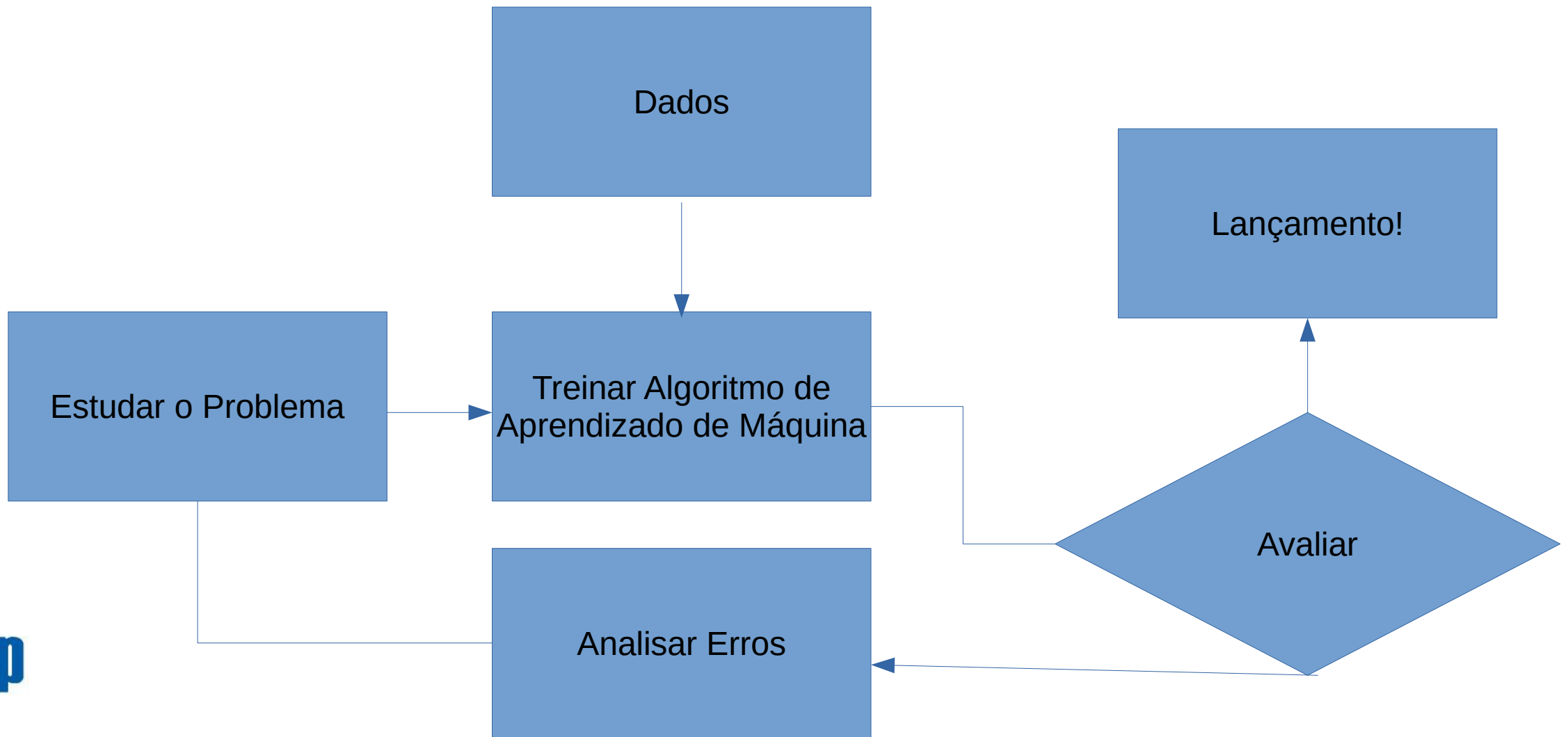


A abordagem tradicional

A abordagem tradicional, seu programa:

- Será uma lista de regras complexas – um programa longo
- Será difícil de manter
- Precisar ser atualizado para se adaptar a novas técnicas.

A abordagem do Aprendizado de Máquina



A abordagem do Aprendizado de Máquina

Na abordagem do Aprendizado de Máquina:

- O filtro detecta mudanças automaticamente nas novas mensagens de Spam
- O programa é muito mais simples e menor pois ele mesmo infere as regras pelos dados
- Como é menor, é mais fácil de manter

A abordagem do Aprendizado de Máquina

Outra área em que o Aprendizado de Máquina vai brilhar é para problemas muito complexos para programas tradicionais ou não tem um algoritmo conhecido.

- Reconhecimento de fala
- Reconhecimento óptico de caracteres (OCR)

A abordagem do Aprendizado de Máquina

Finalmente, Aprendizado de Máquina pode nos ajudar a entender o problema. Podemos estudar o modelo para ter insights.

- Por exemplo: Uma vez que o filtro de Spam tenha sido treinado, podemos extrair as palavras e combinações de palavras que preveem que o e-mail é Spam

A abordagem do Aprendizado de Máquina

Aplicar técnicas de Aprendizagem de Máquina para cavar em grandes quantidades de dados pode nos ajudar a descobrir padrões que não eram imediatamente aparentes.

- Essa abordagem é chamada **Mineração de Dados**

A abordagem do Aprendizado de Máquina

Em resumo, Aprendizado de Máquina é ótimo para:

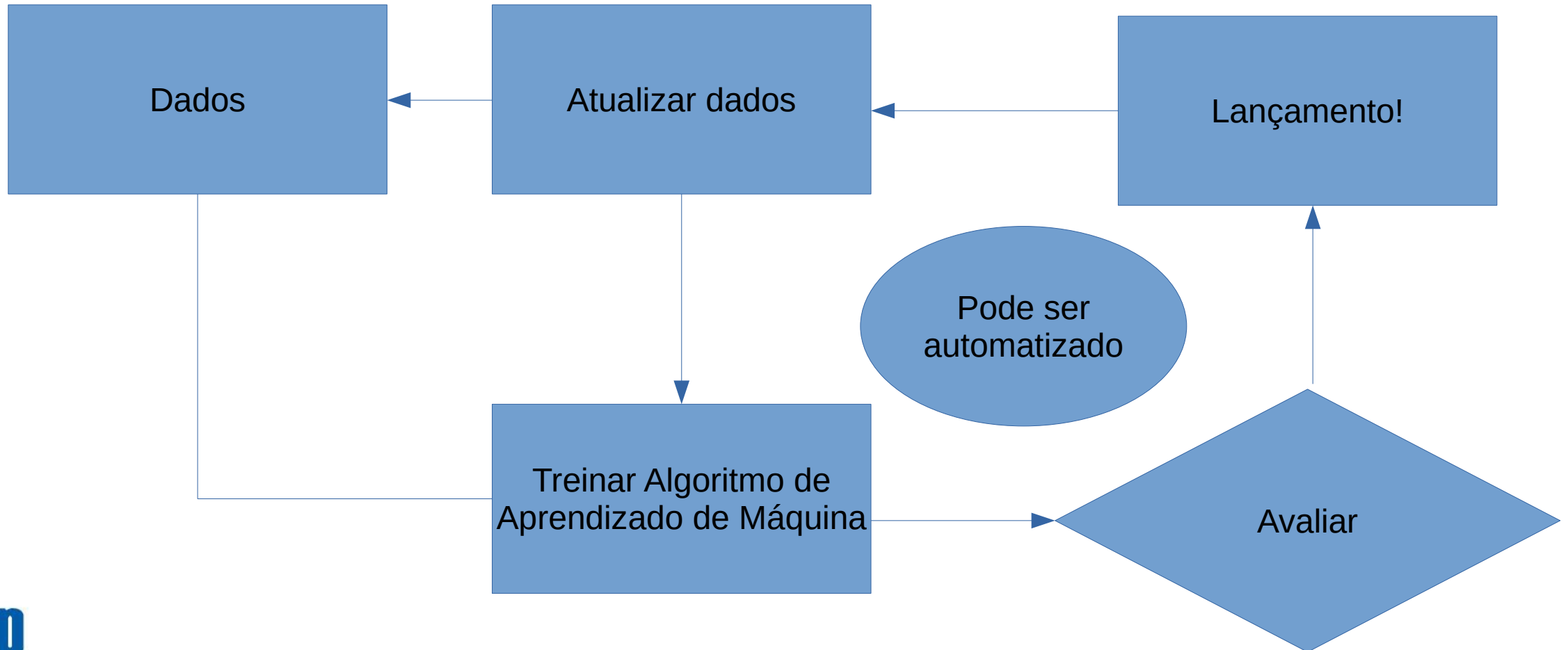
- Problemas para os quais soluções tradicionais exigem muitos ajustes manuais ou longas listas de regras
- Problemas complexos para os quais não existem soluções adequadas com a abordagem tradicional.

A abordagem do Aprendizado de Máquina

Em resumo, Aprendizado de Máquina é ótimo para:

- Ambientes com grande flutuação, que exigem adaptação do algoritmo
- Obter insights sobre problemas complexos e grandes quantidades de dados

Se adaptando a mudança de forma automatizada



Tipos de Sistemas de Aprendizizado de Máquina

Tipos de Sistemas de Aprendizado de Máquina

Vamos classificar em categorias baseadas em se:

- O sistema é treinado com supervisão humana
- Pode aprender incrementalmente a partir de um fluxo de dados
- Funciona a partir da comparação simples com valores conhecidos ou a partir de um modelo

Aprendizado Supervisionado

Aprendizado Supervisionado

Em aprendizado supervisionado, os dados de treinamento possuem as soluções desejadas, chamadas rótulos (labels).

- Por exemplo: o conjunto de dados contém e-mails Spam e normais, marcados de acordo

Essa é uma tarefa de **Classificação**.

Aprendizado Supervisionado

Outra tarefa típica do aprendizado supervisionado é a **Regressão**.

Aprendizado Supervisionado

Regressão é a tarefa de prever um valor numérico alvo, como o preço de um carro, a partir de um conjunto de características chamadas preditores.

- Para treinar o sistema, você precisa de muitos exemplos de carros, incluindo seus preditores e rótulos (preços)

Aprendizado Supervisionado

Note que alguns algoritmos de regressão podem ser utilizados para classificação e vice versa.

- Por exemplo, Regressão Logística é utilizada comumente para classificação pois seu output é a probabilidade de um elemento pertencer a uma classe

Aprendizado Supervisionado

Esses são alguns algoritmos de aprendizado supervisionado:

- k-Nearest Neighbors
- Regressão Linear
- Regressão Logística
- Support Vector Machines (SVM)

Aprendizado Supervisionado

Esses são alguns algoritmos de aprendizado supervisionado:

- Decision Trees
- Random Forests
- Neural Networks

Aprendizado Não Supervisionado

Aprendizado Não Supervisionado

Em aprendizado não supervisionado, o conjunto de dados não está rotulado (unlabeled).

- Por exemplo: Você tem muitos dados sobre os visitantes do seu Blog. Talvez você queira agrupar seus usuários por interesse para realizar marketing direcionado



INSTITUTO BRASILEIRO DE ENSINO,
DESENVOLVIMENTO E PESQUISA