

Aprendizagem de Máquina

Advanced Institute for Artificial Intelligence

<https://advancedinstitute.ai>

Agenda

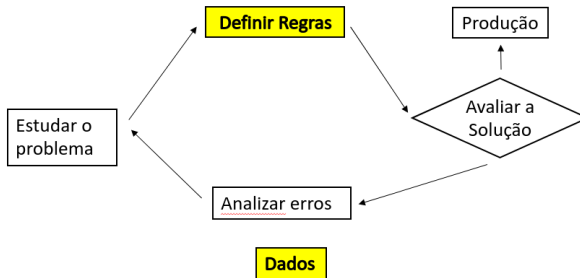
- ☐ Aprendizagem de Máquina
- ☐ Categorias de algoritmos de Aprendizagem de Máquina
- ☐ Introdução a biblioteca Scikit-Learn

- Machine Learning é a ciência (arte) da programação de computadores para que eles possam aprender com os dados.
- "[Machine Learning é o] campo de estudo que oferece aos computadores a capacidade de aprender sem serem explicitamente programados". Arthur Samuel, 1959
- Exemplos de aplicações: identificação de SPAM (email), reconhecimento de objetos em fotos, etc

Sistemas baseados em regras:

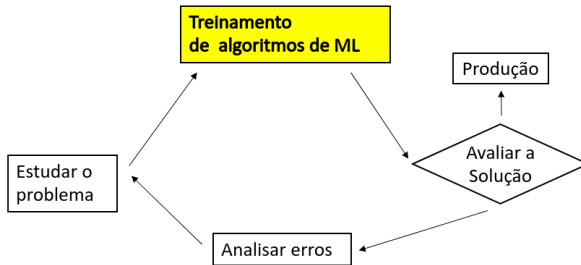
- Natureza dinâmica dos problemas exige redefinição constante das regras
- Sistema de detecção de SPAM de email
 - Spammers identificam que as regras não detectam números e trocam "Dois" por 2
 - Cada pequena mudança exigirá uma adaptação de regras

Aprendizagem de Máquina



- ❑ Um programa tradicional necessitará de uma longa lista de regras
- ❑ Um filtro de spam baseado em aprendizagem de máquina é capaz de utilizar critérios diversos para fazer tal classificação
- ❑ Caracterização de um SPAM pode ser adaptada dinamicamente de acordo com marcações atribuídas pelos usuários

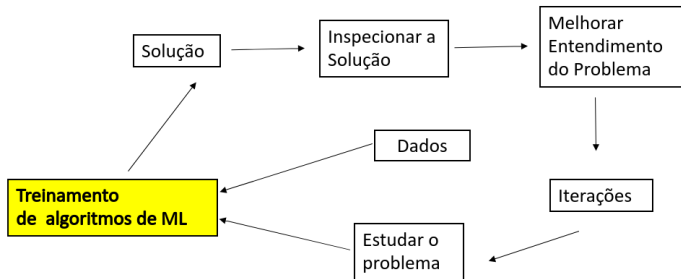
Aprendizagem de Máquina



Processo de utilização de aprendizagem de máquina

- ❑ A utilização de algoritmos de aprendizagem de máquina podem ser melhorados, a partir da análise dos resultados dos algoritmos
- ❑ A aplicação de técnicas de ML para avaliar grandes quantidades de dados pode ajudar a descobrir padrões que não eram aparentes.
- ❑ A utilização de aprendizagem de máquina pode ser entendida como um processo iterativo, em busca de soluções a partir dos dados, e otimização do uso dos dados e algoritmos
- ❑ Esse processo pode ser automatizado

Aprendizagem de Máquina



- Fundamentalmente, o aprendizado de máquina envolve a construção de modelos matemáticos para ajudar entender dados
- Ajustes de parâmetros nos modelos permite que os modelos sejam adaptados aos dados observados
- Desta forma, tais modelo podem ser usados para prever e entender aspectos de dados desconhecidos

Categorias elementares de Algoritmos de Aprendizagem de Máquina

- ☐ Supervisionada
 - Classificação
 - Regressão
- ☐ Não-Supervisionada
 - Agrupamento
 - Redução de Dimensionalidade
- ☐ Semi-Supervisionada

Aprendizagem Supervisionada

- Envolve modelar a relação entre medidas características dos dados e algum rótulo associado aos dados
- O modelo determinado pode ser usado para aplicar rótulos a novos dados
- Tipos de algoritmos supervisionados
 - Classificação: rótulos são categorias discretas
 - Exemplo filtro de spam: Emails são marcados como spam ou não-spam. Modelo classifica novos emails
 - Regressão: rótulos são quantidades contínuas
 - Exemplo: previsão do preço de um carro considerando um conjunto de variáveis preditoras (quilometragem, idade, marca)

No aprendizado não supervisionado os dados de treinamento não são rotulados. O sistema tenta aprender sem referência ou dados anotados.

- ❑ Com base em dados sobre os visitantes de um site. Executar um algoritmo para tentar detectar grupos de visitantes semelhantes.
- ❑ Em nenhum momento você diz ao algoritmo a qual grupo um visitante pertence: ele encontra essas conexões sem ajuda.

Para Qualquer problema a ser investigando como aprendizagem de máquina temos alguns características comuns

- ❑ Amostras (*Samples*): linhas na base de dados
- ❑ Características (*Features*): colunas na base de dados
- ❑ Matriz de Características: Combinação de linhas e características
- ❑ Matriz alvo: coluna que se deseja prever

- ❑ Algoritmos de aprendizagem de máquina normalmente necessitam de uma grande quantidade de dados para apresentar uma solução satisfatória
- ❑ Dados precisam ser representativos em relação ao problema que está sendo investigado
- ❑ Considerar a influência das categorias em relação a base completa
- ❑ Qualidade dos Dados:
 - Considerar detectar e se possível eliminar Outliers e Ruídos