
Machine Learning

4/2/2024
Me. Fernando Nemec

Apresentação

Meu nome é **Fernando Nemec**. Sou formado em Engenharia da Computação. Tenho MBA em Gestão de TI pela FGV e sou Mestre (finalmente!!!) em Finanças pela Fecap.

Trabalho com Tecnologia desde 1998 em funções variadas como programador, coordenador de projetos, gerente de engenharia, diretor de tecnologia.

Atualmente possuo uma empresa de construção e trabalho como desenvolvedor full-stack como consultor.

Se tudo der certo, este ano começo meu doutorado! \o/

Ementa

Calma que vou explicar um por um!

Modelagem Preditiva e Regressão Linear Múltipla, Regressão Logística;
KDD (Knowledge Discovery in Databases), Processos de Descobertas do Conhecimento;
Modelos supervisionados e não supervisionados;
Árvores de Decisão (Random Forest) , SVM – Support Vector Machine, PageRank, Apriori;
Análises de Séries Temporais com Python;
Validação cruzada, Regressão Penalizada (Ridge e Lasso);
OLSR e Ensemble Methods, Adaboost;
Análise Multivariada de Dados Análise de componentes principais com Python;
Análise Multivariada de Dados Análise de Cluster com Python;
Análise de Conglomerados Cluster Analysis (Knn Nearest Neighbor, Kmeans, HClust Cluster Hierárquico);
Inteligência Artificial aplicada a dados;
NLP e Análise de Sentimentos;
Medidas e Métricas de Avaliação;
Praticando modelos de Machine Learning
Processo de Tomada de Decisão E Metodologia de Pesquisa Científica Quantitativa;
Análise descritiva com Python;

Bibliografia

1. Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow (O'Reilly), Aurélien Géron
2. Análise Prática de Séries Temporais (O'Reilly), Aileen Nielson
3. Data Science do Zero (O'Reilly), Joel Grus

O que é aprendizado de máquina?

É o campo de estudo que possibilita aos computadores a habilidade de aprender sem explicitamente programá-los.

- Arthur Samuel (1901-1990), **1959**

Famoso por desenvolver um jogo de damas usando técnicas de IA. Repito: em **1959**.



Por que usar aprendizado de máquina

Porque existem classes de problemas computacionais que os algoritmos de aprendizado de máquina são excelentes.

Exemplos de problemas ótimos para aprendizado de máquina:

- Extensa variabilidade de regras ou adaptabilidade de dados.
- Problemas complexos com grande volume de dados
- Problemas sem solução tradicional aderente

Aplicações

- Detecção de tumores em imagens médicas usando redes neurais convolucionais (CNN);
- Classificação de artigos de notícias utilizando processamento de linguagem natural (PLN);
- Previsão de faturamento da empresa utilizando regressões lineares;
- Fraudes bancárias usando técnicas de detecção de anomalias;
- Segmentação de clientes usando análise de cluster (clusterização);
- Inteligência para um jogo de computador, utilizando aprendizagem por reforço;

Tipos de sistemas de IA

Quanto a supervisão: supervisionado, não supervisionado, semi supervisionado ou por reforço;

Tempo: aprendizado em batch ou em tempo real;

Forma: aprendizado por modelo ou por instância;

Limitações

A principal limitação hoje para o aprendizado de máquina é a disponibilidade de dados.

Para praticamente qualquer algoritmo, será necessário obter dados de treinamento e dados de controle.

E os dados de treinamento precisam ser **representativos** e precisam ter **boa qualidade**.

Características irrelevantes

Outra dificuldade é determinar que dados são relevantes para o modelo, excluindo vieses e informações que pouco contribuem para a assertividade do modelo escolhido.

Sobreaajuste

O sobreajuste acontece quando o modelo, devido a sua complexidade, começa a ser demasiadamente afetado por ruídos.

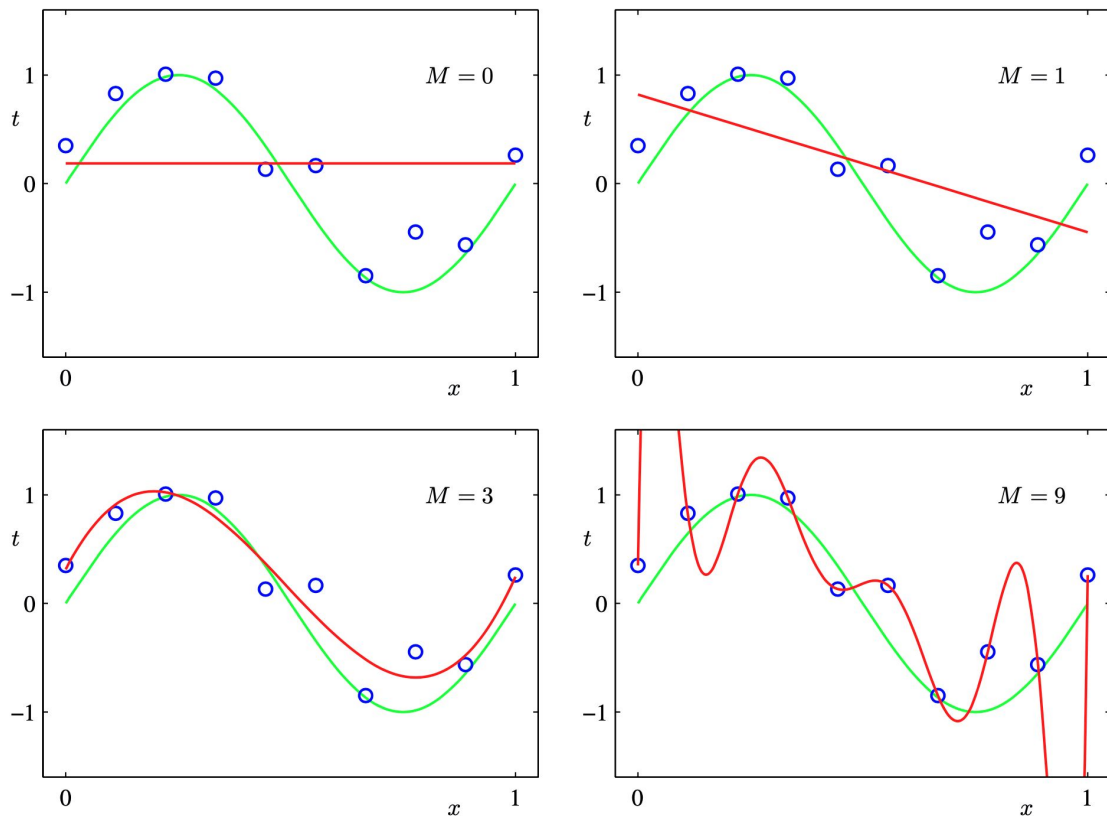


Figure 1.4 Plots of polynomials having various orders M , shown as red curves, fitted to the data set shown in Figure 1.2.

O que é cada coisa na ementa?

Modelagem Preditiva e Regressão Linear Múltipla, Regressão Logística;

Uma das técnicas mais utilizadas na previsão de resultados. A regressão linear pode ser simples ou múltipla. Em algumas situações, que vamos detalhar oportunamente, é necessário ajustes nos preditores para obter uma resposta melhor da regressão.

O que é cada coisa na ementa?

KDD (Knowledge Discovery in Databases), Processos de Descobertas do Conhecimento

Série de processos para obter, selecionar, tratar e descobrir informações novas em bancos de dados relacionais.

O que é cada coisa na ementa?

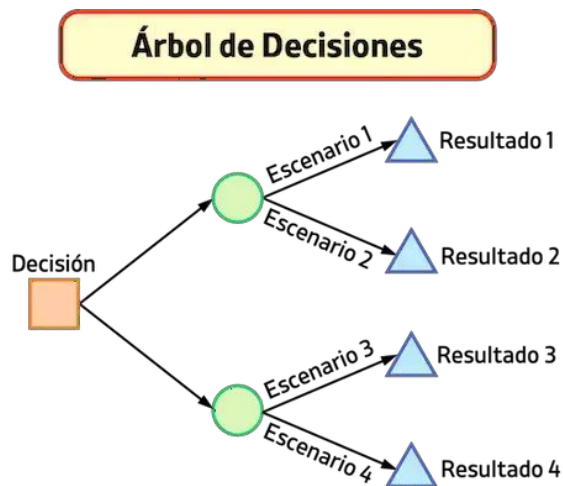
Modelos supervisionados e não supervisionados;

Definição básica de como funcionam modelos supervisionados e não supervisionados e quando é melhor utilizar um ou outro.

O que é cada coisa na ementa?

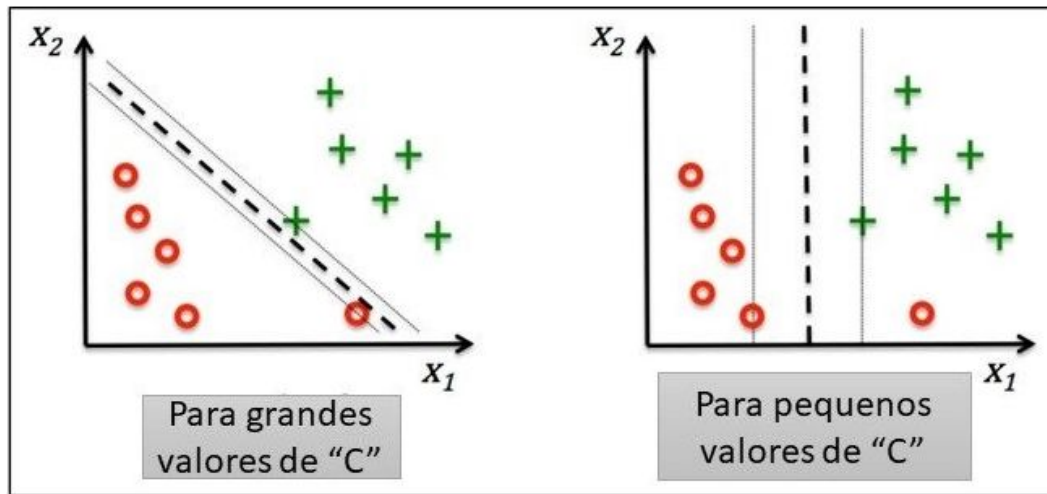
Árvores de Decisão (Random Forest) , SVM – Support Vector Machine, PageRank, Apriori;

Árvore de decisão consiste em um diagrama para visualizar as possíveis escolhas de um determinado algoritmo.



O que é cada coisa na ementa?

SVM são algoritmos de classificação com margem ajustada, como na imagem:



O que é cada coisa na ementa?

Análises de Séries Temporais com Python;

Verificar o comportamento dos dados através do tempo, em busca de padrões como sazonalidade e tendência. Muito usado em Finanças.



O que é cada coisa na ementa?

Validação cruzada, Regressão Penalizada (Ridge e Lasso);

Em algumas situações, para ter um melhor *fit* para um modelo, ajustes são necessários. Ridge e Lasso são ajustes possíveis em uma regressão.

Já a validação cruzada é uma forma de testar a capacidade de generalização do modelo, particionando os dados de treinamento e comparando os resultados obtidos.

O que é cada coisa na ementa?

Ensemble Methods, Adaboost;

Trata da agregação de vários modelos para obtenção de um único modelo, com uma resposta melhor, embora computacionalmente seja muito mais custoso. **AdaBoost** é um tipo de algoritmo de agregação, mas não o único.

O que é cada coisa na ementa?

Análise de componentes principais

Trata da simplificação da dimensionalidade dos dados existentes. Utilizado quando o conjunto de dados existente é grande demais e pode conter redundâncias que somente prejudicam a performance do algoritmo.

O que é cada coisa na ementa?

NLP e Análise de Sentimentos;

Processamento de linguagem natural é a transformação de texto escrito em algo que possa ser compreensível e utilizado pelo seu algoritmo.