Tópicos Avançados em Inteligência Computacional 2 - 2024-1

Ricardo Prudêncio

Responsible AI

- Sistemas de AI são largamente usados
- Vários desses sistemas podem impactar humanos de diferentes formas (e.g., saúde, segurança, financeira, jurídica, política, marketing)
- Os sistemas são complexos
- Questões éticas e de segurança

Responsible AI

Privacy

Explainability

Reliability

Fairness

Trust

Accountability

Disciplina

- Tópicos: ver cronograma
  - Aprendizagem de Máquina (Preditiva)
  - Responsible AI
    - Explicabilidade
    - Confiabilidade
    - Equidade

Disciplina

- Avaliação
  - Participação individual (peso 2) e listas de exercício individual (peso 4) e um projeto final (peso 4)
- Ferramentas e linguagens livres
- Material de Estudo - nos slides e notebooks

Aprendizagem de Máquina

- Sub-área de IA que desenvolve sistemas que melhoram seu desempenho com a experiência

- Abordagens:
  - Data-driven: algoritmos encontram regularidades em dados
  - Classificação, reconhecimento, agrupamento, regressão,...

Aprendizagem de Máquina - Abordagem Data-Driven

- Modelos descritivos
  - Descrevem ou sumarizam dados
- Modelos preditivos
  - Realizam previsões sobre os dados
- Modelos generativos
  - Geram novos dados (e.g., textos, imagens)

Modelos Preditivos
- Classificação

Classificação

- Associar objetos a uma categoria ou classe
  - E.g., diagnóstico de pacientes, classificação risco de um cliente, classificação de documentos,…
- Classificação é feita com base nos atributos dos objetos
  - E.g., diagnóstico de um paciente é feito com base nos sintomas observados e exames realizados
- Para que: alguma tomada de decisão

Exemplos

- Diagnóstico médico
  - Dado um paciente, qual o diagnóstico de uma determinada doença?
  - Qual a criticidade, qual a evolução?
  - Qual a chance de reincidência?

Exemplos

- Classificação de imagens
  - Biometrica (faces, digitais, assinaturas)

Exemplos

- Detecção de anomalias
  - Que transações de crédito são fraudes?
  - Quais equipamentos vão falhar?

Exemplos

- Categorização de usuários em redes sociais
  - Que usuários do Instagram se interessam por roupas? Ou automóveis?
  - Qual o viés político de um usuário no Twitter?

Exemplos

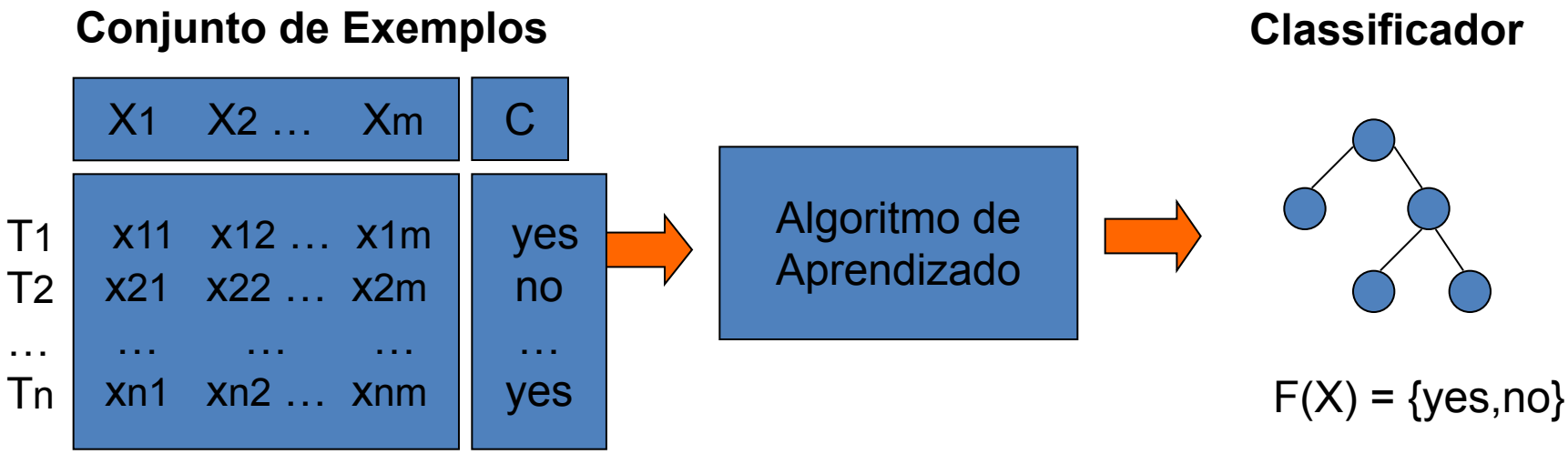
- Análise de sentimentos
  - Esse tweet expressa raiva, alegria, angústia?
  - Essa pessoa está feliz?

Exemplos

- Funções de score
  - Quem será um bom pagador?
  - Qual o risco de sonegação?
  - Quem contratar?

Classificação com AM

- Algoritmo de aprendizagem supervisionada adquire conhecimento a partir de um conjunto de exemplos

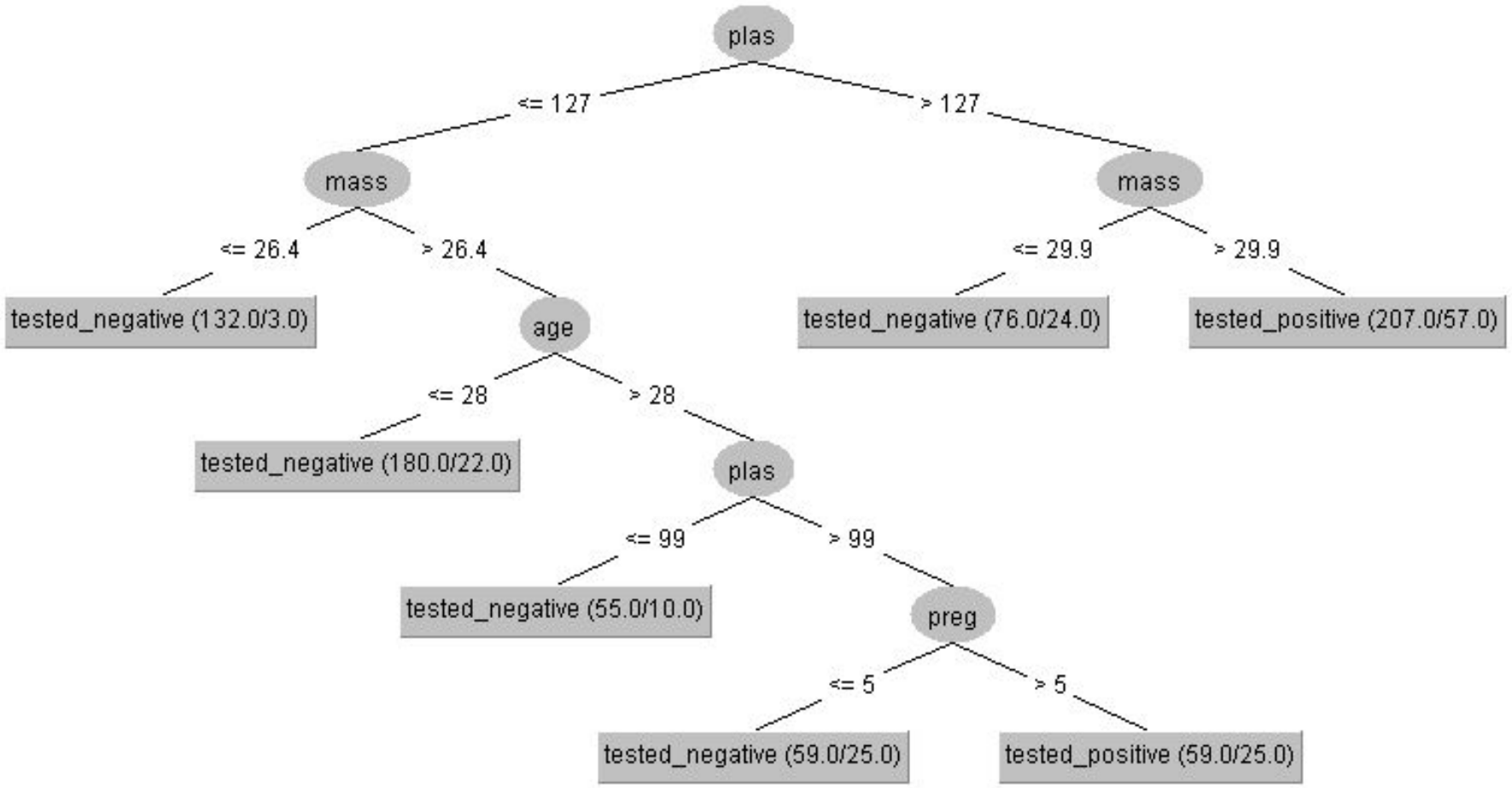


Exemplo

Conjunto de Dados - Diabetes

1: preg Numeric	2: plas Numeric	3: pres Numeric	4: mass Numeric	5: age Numeric	6: **class** Nominal
6.0	148.0	72.0	33.6	50.0	tested_positive
1.0	85.0	66.0	26.6	31.0	tested_negative
8.0	183.0	64.0	23.3	32.0	tested_positive
1.0	89.0	66.0	28.1	21.0	tested_negative
0.0	137.0	40.0	43.1	33.0	tested_positive
5.0	116.0	74.0	25.6	30.0	tested_negative
3.0	78.0	50.0	31.0	26.0	tested_positive
10.0	115.0	0.0	35.3	29.0	tested_negative
2.0	197.0	70.0	30.5	53.0	tested_positive
8.0	125.0	96.0	0.0	54.0	tested_positive
4.0	110.0	92.0	37.6	30.0	tested_negative
10.0	168.0	74.0	38.0	34.0	tested_positive
10.0	139.0	80.0	27.1	57.0	tested_negative
1.0	189.0	60.0	30.1	59.0	tested_positive
5.0	166.0	72.0	25.8	51.0	tested_positive
7.0	100.0	0.0	30.0	32.0	tested_positive
0.0	118.0	84.0	45.8	31.0	tested_positive
7.0	107.0	74.0	29.6	31.0	tested_positive
1.0	103.0	30.0	43.3	33.0	tested_negative
1.0	115.0	70.0	34.6	32.0	tested_positive
3.0	126.0	88.0	39.3	27.0	tested_negative
8.0	99.0	84.0	35.4	50.0	tested_negative
7.0	196.0	90.0	39.8	41.0	tested_positive
9.0	119.0	80.0	29.0	29.0	tested_positive
11.0	143.0	94.0	36.6	51.0	tested_positive
10.0	125.0	70.0	31.1	41.0	tested_positive
7.0	147.0	76.0	39.4	43.0	tested_positive
1.0	97.0	66.0	23.2	22.0	tested_negative
13.0	145.0	82.0	22.2	57.0	tested_negative

Exemplo - Modelo de Árvore de Decisão



Exemplo - Modelo de Regressão Logística

Variable	Class tested_positive
preg	0.1188
plas	0.0338
pres	- 0.0135
mass	0.091
age	0.0171

$$P(Y=1 \mid x_1,...,x_p) = \frac{1}{1+\exp(-(\beta+\alpha_1 x_1+...+\alpha_p x_p))}$$

Classificação - Definições

- Exemplo (ou instância)
  - Tupla com atributos que descrevem um objeto de interesse + classe do exemplo
    - E.g., dados de um paciente + doença
    - E.g., medidas de complexidade de software + {bug ou não bug}
- Atributos Preditores
  - Característica de um exemplo usada para classificação
- Atributo Alvo
  - Problemas de classificação binários ou multi-class

Classificação - Definições

- Conjunto de Treinamento
  - Coletado da base de dados e etiquetado (rolutado) geralmente por um humano
  - Usado para construir um classificador
- Conjunto de Teste
  - Conjunto usado para avaliar a qualidade do classificador gerado
- Classificador (Modelo)
  - Resultado retornado pelo indutor (aproxima a função real de classificação)

Quais os desafios?

- Dados
  - Dados enviesados, subrepresentados, com mudanças de distribuição ao longo do tempo, classes desbalanceadas, qualidade das variáveis
- Modelos complexos
  - Dificuldades de interpretação, auditoria, manutenção, adaptação, etc

Referências

Kaur D. et al. (2022) Trustworthy Artificial Intelligence: A Review. ACM Computing Surveys.

Flach P. (2012). Machine Learning: The Art and Science of Algorithms that Make Sense of Data