

Machine learning

Componentes do aprendizado

I) Entrada (x)

Dado do cliente \rightarrow vetor com atributos \rightarrow features

II) Saída (y)

Rótulo (alvo) \rightarrow Bom ou mau pagador?

III) Função objetivo ($f: x \rightarrow y$)

função ideal (f^*)

IV) Dados

Registros históricos rotulados

\rightarrow representam um cliente anterior

V) Hipótese ($g: x \rightarrow y$)

- função a ser utilizada
- função ideal que associaria qualquer entrada x à saída y
- É desconhecida, justamente o que tentamos aproximar usando ML
- O aprendizado tenta estimar uma função com base nos dados rotulados

$$f^*: \mathcal{X} \rightarrow \{0,1\}$$

$$f(x) \approx f^*(x)$$

tipos de aprendizado

Lazy: IBL - instance-based learning

Eager: MBL - model-based learning

Lazy

- Não generaliza e nem constrói um modelo durante o treinamento. Em vez disso, ele armazena os dados e só faz cálculos quando precisa fazer a predição

\rightarrow Características

- Predição é lenta, pois o modelo aprende na hora de responder
- Usa instâncias dos dados diretamente para tomar decisões
- Não tenta aprender uma função explícita

\rightarrow Vantagens

- Simples e fácil de implementar
- Adapta-se bem a mudanças nos dados
- Não precisa de suposições fortes sobre os dados

\rightarrow Desvantagens

- Alto custo computacional na predição
- Pouca capacidade de generalização
- Pode ser sensível a ruídos e a dados irrelevantes

Eager

- O modelo tenta generalizar a partir de dados de treinamento, aprendendo uma função explícita antes de fazer qualquer predição

\rightarrow Características

- O treinamento é mais demorado, pois envolve ajuste de parâmetros ou otimização
- Predição é rápida, pois o modelo já tem uma função pronta
- tenta adaptar padrões globais aos dados

\rightarrow Vantagens

- Predições rápidas e eficientes
- Melhor capacidade de generalização
- Reduz o espaço de armazenamento

\rightarrow Desvantagens

- O treinamento pode ser caro computacionalmente
- Pode sofrer com overfitting ou underfitting se mal ajustado
- Menos flexível a novos dados não vistos