

INTELIGÊNCIA ARTIFICIAL

Hierarquia do aprendizado de máquina:

- Aprendizado Indutivo:
 - Aprendizado supervisionado
 - Classificação
 - Regressão
 - Aprendizado não-supervisionado

Aprendizado indutivo: permite obter conclusões genéricas sobre um conjunto particular de exemplos

Caracterizado pelo raciocínio originado em um conceito específico, que é ao longo do processo, generalizado

Utilizado para derivar conhecimento novo e prever eventos futuros

Efetuada a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema

Aprendizado supervisionado: há um conjunto de dados rotulado

Alguns métodos:

- classificação e regressão
 - * redes neurais perceptron multicamadas (MLP)
- Classificação:
 - * árvores de decisão
 - * Naive Bayes (aprendizado probabilístico)
 - * Redes Neurais Perceptron (uma única camada)

Aprendizado não-supervisionado: ignora os rótulos e agrupa os elementos. Verifica as classes dos produtos que podem ser agrupados de alguma maneira, formando agrupamentos ou clusters.

Após a determinação dos agrupamentos, em geral, é necessário uma análise para determinar o que cada agrupamento significa no contexto do problema sendo analisado

Fases de aprendizado

1 Dados

2 Pré-processamento (há um impacto direto no algoritmo de aprendizado)

- * Seleciona atributos
- * Eliminar ruídos
- * Tratar valores perdidos

3 Processamento (AM – Aprendizado de Máquina)

4 Pós-processamento

Conjunto de treinamento e teste

De modo geral, o conjunto de exemplos disponíveis deve ser dividido em dois conjuntos disjuntos:

- conjunto de treinamento:

- * usado para aprendizado do conceito (construção do modelo)

- conjunto de teste:

- * usado para medir o desempenho e grau de generalização do modelo construído

- * a disjunção dos conjuntos é exigida para tornar essa medida estaticamente válida

-> O algoritmo será bom se ele for capaz de analisar algo que o programador não viu

Classificador

Dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador

-> prever rótulos

-> não com maior precisão possível, mas com maior accuracy

Formalmente, um exemplo pode ser representado pelo par:

$(x, y) = (x, f(x))$,

Em que:

- * X é a entrada

- * $F(x)$ é a saída (rótulo)

- * Indução ou inferência indutiva: dada uma coleção de exemplos de f , retornar uma função h que aproxime f

- * H é denominada uma hipótese sobre f

-> objetivo de achar uma função h que tenha um bom acerto e seja o mais simples possível. Não pode ser muito complexo, pois isso significa que o modelo decorou os dados, por isso deve ser simples.

Erro e acurácia

Erro: para cada exemplo passado na fase de teste compara-se com a hipótese gerada pelo modelo. Se for diferente, significa que houve um erro

$$\text{Err}(h) = 1/n (\text{somatória}) | | y_i \neq h(x_i) | |$$

O operador $| | E | |$ retorna:

1 se verdadeiro e 0 se falso

$$\text{Acc}(h) = 1 - \text{err}(h)$$

Regressão: distância entre o valor real e o predito

$$\text{Mse} - \text{err}(h) = 1/n (\text{somatória}) (y_i - h(x_i))^2$$

$$\text{Mad} - \text{err}(h) = 1/n (\text{somatória}) | y_i - h(x_i) |$$

Mse: erro quadrático médio (*mean squared error*)

Mad: distância absoluta média (*mean absolute distance*)

Distribuição de Classes

Classe majoritária possui a maior distribuição em termos de porcentagem

- Erro majoritário:

Dada a distribuição de k classes em um conjunto de exemplos T , pode-se calcular o erro majoritário desse conjunto como: $\text{Maj} - \text{err } T = 1 - \text{Max Distr } C$

É independente do algoritmo de aprendizado.

Fornecer um limiar máximo abaixo do qual o erro de um classificador deve ficar.

Prevalência de Classe: quando há um desbalanceamento de classes no conjunto de exemplos.

-> uma hipótese boa é uma hipótese simples, mas que dê para distinguir um conceito do outro.

-> Não pode ser complexo, pois não conseguirá generalizar.

Overfitting

Ocorre quando a hipótese extraída a partir dos dados é muito específica para o conjunto de treinamento

A hipótese apresenta um bom desempenho para o conjunto de treinamento, mas um desempenho ruim para os casos fora desse conjunto

-> avaliação de modelo realizado acima do conjunto de teste, pois a porcentagem de erro pode ser diferente entre o conjunto de treinamento e de teste.

Underfitting

A hipótese apresenta um desempenho ruim tanto no conjunto de treinamento como de teste.

- poucos exemplos representados foram dados ao sistema de aprendizado

- o usuário predefiniu um tamanho muito pequeno para o classificador

Métodos para avaliação de Classificadores

- Métodos de amostragem: serve para que possa ser implementada as informações básicas para o aprendizado. Utilizado para dividir o conjunto de dados.

* holdout: método antigo. Usado somente em análises exploratórias

* amostragem aleatória:

r-fold cross-validation: divide o conjunto em folds. Há um cálculo em várias rodadas, sendo que em cada uma um fold é deixado como aprendizado e o restante em modo teste. Em cada rodada também é armazenada a taxa de erro.

r-fold stratified cross-validation: erro médio do classificador um grau de confiança.

leave-one-out

bootstrap

Amostragem (resampling): metodologia de avaliação, utilizada para comparar dois algoritmos de classificação

Para estimar o erro a amostra deve ser aleatoriamente escolhida e não deve ser pré-selecionada

Para problemas reais, tem-se uma amostra de uma única população, de tamanho n , e a tarefa é estimar o erro verdadeiro para essa população

-> lws (análise exploratória): criado por Fisher

- Matriz de confusão: oferece uma medida da eficácia do classificador, mostrando o número de classificações corretas versus as classificações previstas para cada classe.

Para 2 classes

Medidas de erro:

Taxa de FP = $FP / (FP + VN)$ //FP: falso positivo

Taxa de FN = $FN / (VP + FN)$ //falso negativo

- Estimando a taxa de erro de um classificador:

150 exemplos no fold

Armazenar os erros em cada fold e dividir pelo número total: 15 exemplos /fold

Calcula a taxa de erro de todos, soma tudo e divide por 150

Árvores de decisão

Algoritmo ID3 – robusto.

Conjunto de regras disjuntivas: conjunto de regras unidas por disjunções (ou)

Conhecimento expresso baseado em regras disjuntivas (conjunto de hipóteses com disjunções, ou seja, caminho da árvore)

Hipóteses

É uma aproximação da função representada por aquele conhecimento, que divide o espaço.

Cada modelo de aprendizado de máquina, gera uma hipótese.

O que define a hipótese é o paradigma de funções, disjunção de regras.

A hipótese deve ser boa o suficiente, mas simples.