

Projeto Aprendizado de Máquina
CTC- 17 Inteligência Artificial
(Trabalho Individual ou em Grupo de dois ou três alunos)
Prof. Paulo André Castro

1. Objetivo

Exercitar e fixar conhecimentos adquiridos sobre Aprendizado de Máquina utilizando árvores de decisão e/ou redes bayesianas ingênuas (*Naïve Bayes*) utilizando uma base de dados fornecida.

2. Descrição do Trabalho

2.1. Base de dados (dataset)

Neste projeto, devem ser usadas as informações de classificações de filmes (1 a cinco estrelas) para os filmes e usuários fornecidos no pacote de dados (ml-1m.zip) obtidas por MovieLen (<http://grouplens.org/datasets/movielens/>) e disponibilizada no site da disciplina. A base de dados ml-1m oferece informações sobre usuários (UserID::Gender::Age::Occupation::Zip-code), filmes (MovieID::Title::Genres) e as classificações dadas, aproximadamente 1 milhão de classificações dadas por 6000 usuários para 4000 diferentes filmes. Veja o arquivo README para mais detalhes.

2.2. Classificador baseado em árvore de decisão (ou Naive Bayes)

Utilizando a base de dados fornecida, criar um **classificador baseado em árvore de decisão** que classifique um dado filme com base nas informações disponíveis sobre o filme e sobre o usuário. Alternativamente, pode-se criar um classificador baseado na técnica *Naive Bayes*.

2.3. Classificador a priori

Crie um classificador *a priori*, isto é que não usa nenhuma informação além da própria identificação do filme. Faça a média (truncada) das classificações para cada filme.

2.4. Análise Comparativa

Compare os dois classificadores utilizando: taxa de acerto, matriz de confusão, erro quadrático médio e estatística kappa.

Para fazer a comparação, selecione pelo menos dez filmes que você assistiu dentro da base e dê sua classificação em estrelas. Discuta qual classificador entre os dois é melhor e como poderia ser criado um classificador ainda melhor.

3. Material a ser Entregue e Prazo

Material: Relatório e Código

Prazo de Entrega: 28/outubro/2016

Relatório do Projeto (arquivo em formato pdf até 4 páginas) com:

Título: Projeto Aprendizado de Máquina

1.Nome:

2. Resultados Obtidos

2.1. Descrição dos Classificadores

2.2. Dados e Resultados da comparação

2.3. Discussão e sugestão de melhorias para o classificador

3. Conclusões: Comentários e sugestões sobre o trabalho (complexidade/facilidade, sugestões, etc.).

4. Descrição da Implementação: Linguagem e IDE utilizados, outros comentários eventualmente necessários para a execução do projeto.

Código do Projeto

Código-fonte do Sistema (em C, C++, C#, Python ou Java).

Bom Trabalho!
Prof. Paulo André Castro
pauloac@ita.br

PS: Se o trabalho for em grupo, o classificador Naive Bayes passa a ser obrigatório e a classificação deve envolver os três classificadores.