

01

Árvore de decisão

Transcrição

Até o momento mexemos na aba "Preprocess" do Weka. Agora vamos começar a trabalhar com classificação, então clicaremos na aba "Classify".

Nessa aba, a primeira opção que aparece é "Classifier", pois temos que escolher um classificador. Para isso, vamos em "Choose" e na pasta "trees". Nela, escolheremos especificamente o "Random forest". Clicaremos em "Start" e iniciaremos um treinamento para a classificação.

Fazendo isso, poderemos também verificar logo abaixo em "Text options" se é "deposito" que está selecionado, pois "deposito" é nosso alvo de classificação. Queremos classificar os clientes entre os que farão e os que não farão o depósito.

Do lado direito, aparece o "Classifier output", a saída do treinamento do nosso classificador. Essa saída será diferente para cada algoritmo com que formos trabalhar, mas ao final sempre aparecerá um "Summary", resumo com algumas saídas. Nele podemos ver uma porcentagem referente à taxa de acerto, que mostra, por exemplo, que após o treinamento, ele acerta 85% das instâncias que tenta classificar, errando 14% das vezes.

Do lado esquerdo da tela será mostrado o número de instâncias que representam esses os 85% da base de dados que ele testou. Agora, melhoraremos esse resultado. Clicaremos no "Random forest" e pensaremos em como fazer isso a partir das opções oferecidas.

Olhando para as opções, veremos o campo "About", em que aparece do que o algoritmo se trata. Em "More" veremos o que são cada uma das opções listadas nessa janela, se formos explorar cada uma delas. Focaremos no número de iterações ou "numIterations". Uma iteração significa que foi realizada uma classificação e o programa tentou realizar um refinamento desse algoritmo de classificação. Depois disso, haverá ainda uma nova classificação e um novo refinamento. Nesse caso, ele tentará fazer isso por 100 vezes.

Modificaremos esse número de 100 para 200. Será que tentando refinar por mais vezes a classificação será melhorada? Vamos testar. Clicaremos em "Start" novamente e devemos ter essa pergunta em mente, se ele melhorará com a nossa tentativa de refinar por mais iterações, ou se ele vai parar depois de um determinado número por não conseguir refinar mais a classificação.

O Weka demorará um pouco mais para fazer essa classificação e agora teremos uma taxa de acerto diferente, um pouco mais alta do que a anterior. 20 instâncias a mais terão sido classificadas corretamente, indo de 85.7% para 85.95%. Poderemos ver o resultado anterior clicando em "Results list".

Um ponto importante é que se clicarmos no teste anterior, ele não retornará para as opções que utilizamos no teste anterior. Precisaremos ter sempre anotadas as opções que usamos para conseguir os melhores resultados de treinamento.