CITS3401 Data Exploration & Mining Project 2

Classifying Poker hands using Weka

Aleck Greenham 20362627
Ash Tyndall 20915779

27th May 2013

Abstract

This document details the analysis of how different machine learning algorithms handle the classification of different Poker "hands" (collections of 5 playing cards) as one of the nine well-defined classes. Where the analysis requirements [**?**] were ambiguous or incomplete, reasonable assuptions were made and documented.

Introduction

Limitations

Requirements

Assumptions

The following assumptions were made:

- item

1. item

Implementation

Data Restructing

The originally provided data was in CSV format, while Weka only easily supports its own ARFF format. The tool linked to in the design document [**?**] unfortunately could not handle the large testing set (presenting error 500), so a suitable Python replacement was devised.

Furthermore, the initial data produced had all of its attributes described as "numeric" in the produced ARFF format. Noting that the explanation of attributes document [**?**] described the "suit of card" and "class of hand" attribute as "ordinal" rather than "numerical" (as the rank of card was), we elected to modify the data set so that attributes C1 through C5 and CLASS were represented in Weka as a "nominal specification" (the Weka form of an enumeration) instead.

The Python script was thusly modified to map certain numbers to representative letter combinations to assist in the interpretation of the enumeration.

Data Import

References

[1] CITS3401 Data Exploration and Mining - Project 2, http://undergraduate.csse.uwa.edu.au/units/CITS3401/labs/proj2-2013.html

[2] Online CSV to ARFF conversion tool, http://slavnik.fe.uni-lj.si/markot/csv2arff/csv2arff.php

[3] Explanation of Poker Hand data attributes, http://undergraduate.csse.uwa.edu.au/units/CITS3401/labs hand.names

[4] Attribute-Relation File Format, http://www.cs.waikato.ac.nz/ml/weka/arff.html