Poker Rule Induction

Kaggle Competition

Jessica Thomas
Alan Dodge
Spencer McEwan

The Problem

Generate rules for classifying poker hands without making assumptions about the rules of poker.

Hand Classifications:

0: Nothing in hand; not a recognized poker hand

1: One pair; one pair of equal ranks within five cards

2: Two pairs; two pairs of equal ranks within five cards

3: Three of a kind; three equal ranks within five cards

4: Straight; five cards, sequentially ranked with no gaps

5: Flush; five cards with the same suit

6: Full house; pair + different rank three of a kind

7: Four of a kind; four equal ranks within five cards

8: Straight flush; straight + flush

9: Royal flush; {Ace, King, Queen, Jack, Ten} + flush

The Data

The Data

25,010 training hands

1 million testing hands

- These hands are not already classified.

S1	C1	S2	C2	S3	C3	S4	C4	S5	C5	hand
1	4	3	6	1	12	3	11	2	7	0
1	11	4	1	3	7	4	11	2	1	2
4	6	4	11	1	11	2	11	2	6	6

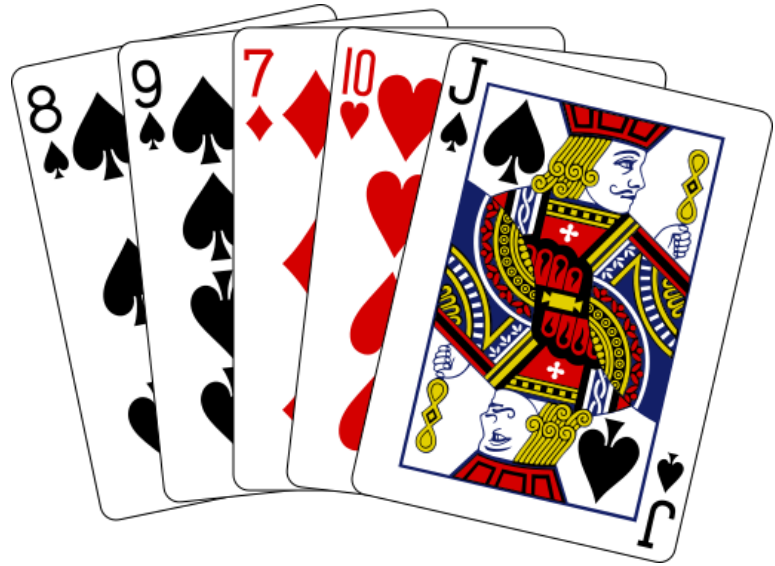| id	S1	C1	S2	C2	S3	C3	S4	C4	S5	C5
1	1	10	2	2	3	3	3	8	1	1
2	2	13	3	5	3	7	4	6	1	4
3	1	3	1	11	2	8	2	1	2	4

C1 = Card 1 S1 = Suit for Card 1, C2 = Card 2, S2 = Suit for Card 2, etc.

'hand' describes the classification (0-9) of the poker hand.

The Data

The input hand:

S1	C1	S2	C2	S3	C3	S4	C4	S5	C5	hand
4	8	2	9	2	7	1	10	4	11	4

Data Mining

Three Steps:

1. Generate

   - Generate rules for the training hands.

2. Generalize

   - Generalize rules based on patterns between hand classifications.

3. Evaluate

   - Evaluated the testing hands.

Generating

Generate rules based on adjacency, rank equivalence, and suits for each hand.

- Generated rules are based on assumptions about a deck of cards: there are suits and cards within a suit have an ordering.

Generating

S1	C1	S2	C2	S3	C3	S4	C4	S5	C5	hand
4	8	4	9	2	7	1	10	4	11	4

Output: {'4': [(1, '!', 2), (1, '!', 3), (1, '!', 4), (1, '!', 5), (2, '!', 3), (2, '!', 4), (2, '!', 5), (3, '!', 4), (3, '!', 5), (4, '!', 5), ('e', 0), ('m', 11), ('n', 7), (1, '#', 2), (1, '%', 3), (1, '%', 4), (1, '#', 5), (2, '%', 3), (2, '%', 4), (2, '#', 5), (3, '%', 4), (3, '%', 5), (4, '%', 5), ('s', 3), (1, '+', 2), (1, '+', 3), (2, '+', 4), (4, '+', 5), ('a', 4), ('S3', 2), ('S2', 4), ('S1', 4), ('S5', 4), ('S4', 1), ('C3', 7), ('C2', 9), ('C1', 8), ('C5', 11), ('C4', 10)]}

Generalizing

Generate general rules for each hand classification.

- Find rules that all hands in a class have in common and eliminate all others.

Evaluating

Check the number of successfully classified hands

-   Ten-fold cross-validation
    -   Train on nine parts of the testing file, and based on the rules generated from those nine parts, we classify the remaining hands and check whether our classification is correct.

Classify the testing data set hands.

-   Generate output for Kaggle Competition.

Results

Using testing data:

- 99.868 % accuracy
    - (when unclassified rules were classified as "not a valid hand")
- 99.9843% coverage

Using training data:

- Not ideal classification if the subset of training data for that class is too small (it might keep erroneous rules about the class).
- Close to 100% accuracy and coverage with ten-fold cross validation
- Difficult to identify the "round" feature of the Ace (main source of misclassification).

28	↓3	Bronswick	0.99994	5	Tue, 17 Mar 2015 08:59:20
29	new	li35y	0.99989	1	Fri, 27 Mar 2015 03:00:35
30	↓4	Jake Teg	0.99974	2	Thu, 12 Mar 2015 19:51:57
31	↓4	Andrew Geisler	0.99958	3	Sun, 11 Jan 2015 18:17:56 (-0.1h)
32	↓4	sna	0.99944	6	Thu, 05 Feb 2015 07:21:33 (-17.7h)
33	↓4	Marco Scattolin	0.99941	1	Mon, 08 Dec 2014 11:49:44
34	↓4	Manish Kumar	0.99941	4	Wed, 07 Jan 2015 04:34:36
35	new	Arthur Pajot	0.99938	3	Wed, 25 Mar 2015 22:41:25
36	↓5	Ryan Knitter	0.99898	3	Tue, 24 Mar 2015 21:21:39 (-33.9h)
37	new	**jurassica** 👤	**0.99868**	**1**	**Sun, 29 Mar 2015 23:13:37**
38	↓6	erickmiranda	0.99858	3	Mon, 09 Mar 2015 13:56:58 (-2.8d)
39	↓6	Bonferroni27	0.99840	1	Sun, 07 Dec 2014 04:37:21
40	↓6	PRB	0.99813	2	Sat, 10 Jan 2015 07:26:42
41	↓6	CrazyHotHot	0.99761	1	Tue, 10 Feb 2015 01:43:08
42	↓6	RLai	0.99747	1	Sun, 18 Jan 2015 19:32:40
43	↓6	Nikhil Mitra	0.99738	2	Wed, 11 Feb 2015 14:00:48
44	↓6	Nathan Patrick Taylor	0.99587	3	Mon, 22 Dec 2014 23:57:07

Kaggle Competition Leaderboard!

Next Steps

- Extend learning to any other type of card game, even uno or crazy eights (anything that can be given a name for a hand).

- Change the classification process to improve coverage of hard-to-identify cases.