Assignment 3

Practical Data Science

In a game of Bridge, there are four players – North, South, East and West. Each player is served a hand of 13 cards. We denote the hand of a player like:

$$K6.Q72.J983.KQT9$$

where each block (e.g. $K6$) denotes cards (e.g. King and 6 in this case) of a specific unspecified suit, for a total of 4 suits (not in order, spades, hearts, diamonds and clubs).

Given South's hand, your task is predict which of the following facts(s) is true for a **given game**:

- Of the number of tricks that can be taken by North/South with South leading, the greatest number of tricks is attained for No Trump (NT)
- Of the number of tricks that can be taken by North/South with South leading, the greatest number of tricks is attained for Spade Trumps (S)
- Of the number of tricks that can be taken by North/South with South leading, the greatest number of tricks is attained for Heart Trumps (H)
- Of the number of tricks that can be taken by North/South with South leading, the greatest number of tricks is attained for Diamond Trumps (D)
- Of the number of tricks that can be taken by North/South with South leading, the greatest number of tricks is attained for Club Trumps (C)

For instance, if the greatest number of tricks is attained for **No Trumps** (NT), **Heart Trumps** (H) and **Club Trumps** (C), the label assigned would be "`1,0,1,0,1`". Note that as the hands of North, East and West are **not** known to you, for the same input hand (i.e. South's hand), there could be multiple possible label sets associated with it (depending on the others' hands).

Your task is to build a classifier that predicts which of the label(s) holds true for our test set by training it on the provided training dataset:

```
"T4.AKJ9.J8642.AT","1,0,1,0,1"
"6.KT7.A53.AQT865","1,0,0,1,1"
"J.943.AQ982.T973","0,0,1,0,0"
```

For each test sample (one per line), produce the corresponding comma-separated list of output labels, e.g. `0,0,1,0,0` without quotation marks). Code your solution in the `a3` directory, and include a bash script, `a3/run.sh input output`, where `input` is the user-specified path to the input file and `output` is the user-specified path to the output file. Do not hard-code any paths to the input or output file.

The training file can be found at `/home1/soham.pal/PDS_data/a3_train.csv`. You are free to hardcode this path if you so wish, or copy it into your home directory. Please do not create a virtual environment inside the `a3` directory. **If you use any online resource, please cite it in your code using a comment. This is an individual assignment, please do not work in teams, copy or share others' code.**