Homework 5 (Psych186B)
Federation Intelligence Exercise (30 points)

Posted on Feb 2; Due 1159pm, Feb 15

The goal of this homework assignment is train a linear associator to identify
the origins of ships based on sensor scans, so that the Enterprise will be able
to take appropriate action when another ship approaches.

Incoming ships, particularly when far away, may generate noisy sensor readings
that sometimes give misleading or partial information. Sometimes only a few
letters of the name of the ship can be retrieved from the automatic transponder,
or the hailing transmitter is not readable, or the color of the ship or its
shape can be only crudely discerned. You should be able to take this partial,
noisy information and (in descending order of importance) tell:

   (1) Whether the incoming ship is liable to be hostile or peaceful.
   (2) Whether the Enterprise should enter a state of heightened alertness.
   (3) The tentative identification of the ship's system of origin.

You will be classifying new ships into four origins Klingon, Romulan, Antarean,
and Federation. Klingons are to be treated as hostile. Romulans require Alert
status. Antareans and the Federation must be considered friendly.

Data from previous encounters provides enough information to characterize new
ships, even in corrupted form.

Archival Intelligence Data Table for Training Neural Network

Name	Planet of origin	Warp Drive Vibration Index (Murds)	Hailing Transponder Freq. (gigaHz)	Surface Reflect. (color)	Ratio of long to short axis	Req. action
Grotz	Klingon	6.9	1006.4	Black	3.5	Hostile
Tlarr	Klingon	7.0	994.3	Black	2.3	Hostile
Tribok	Klingon	7.3	978.1	Dark Gray	2.8	Hostile
Brogut	Klingon	7.1	1005.4	Dark Gray	3.0	Hostile
Glorek	Klingon	7.1	1001.8	Light Gray	1.0	Hostile
Lorif	Romulan	7.3	980.4	Dark Blue	1.6	Alert
Rallev	Romulan	7.4	977.2	Dark Green	1.8	Alert
Willosh	Romulan	7.3	947.9	Light Gray	1.9	Alert
Loshar	Romulan	7.2	955.8	Light Blue	2.1	Alert
Sarash	Romulan	7.4	960.7	Light Gray	2.3	Alert
A2231	Antarean	6.7	1010.9	Pink	1.2	Friendly
E7763	Antarean	6.8	1033.2	Orange	1.2	Friendly
E9091	Antarean	6.5	1025.4	Light Blue	1.1	Friendly
A0199	Antarean	6.8	1066.2	Yellow	1.3	Friendly
A1091	Antarean	6.7	1015.0	Light Blue	1.0	Friendly
Daisy	Federation	6.7	1050.0	White	1.9	Friendly
Rosehip	Federation	6.8	1055.0	Light Gray	2.0	Friendly
Gardenia	Federation	6.5	1045.0	White	2.1	Friendly
Herb	Federation	6.4	1065.0	Light Gray	2.6	Friendly
Cinnamon	Federation	6.5	1055.0	Light Gray	1.7	Friendly

We are assured by Starfleet Command that this is a fair and representative set of ships. We can use them to faithfully represent the navies of their planets of origin. Note though, that the *Glorek* is atypical in some respects. For one thing, it is nearly spherical, much more typical of the peaceful Antareans. In fact, the *Glorek* actually is a captured Antarean ship, the *E3120.*

Starfleet Command claims that the use of the regularities in the data from the 20 known ships lead to correct identification of response and planet of origin of all 20 ships the first time the programs were run. However, it is possible that as few as 18 completely correct answers might be obtained with a good coding and proper programming because of the statistical nature of the samples. Therefore, identifying 18 out of 20 new ships correctly will be considered to be full score for classification.

The testing set provided for the exercise consists of 20 sets of partial information, which the main Enterprise computer has provided for you, to correspond to the kind of noisy data that would be seen in reality.

For each ship, identify its origin (Klingon, Romulan, Antarean, or Federation) and the required action (Hostile, Alert, or Friendly).

When building your linear associator, consider:
- How to represent the different ship features in your input vectors. You will need to convert the strings to some numerical value, but how?
- How to represent the ship classification in your output vectors
- How to deal with the missing data entries (e.g., missing letters in ship name, uncertain ship color, interval for frequency)

When selecting ship features to represent in your input vectors, it may be helpful to look through the noisy data first and see what types of partial information are given. These features can be anything from components of ship names, range or threshold for numerical data (e.g., ratio), to categories of colors.

When defining your f vectors, remember that each dimension represents a feature. You may use binary (0,1), or bipolar (1, 0, -1) coding for each feature, or standardize numerical values.

Also make note that you are NOT allowed to use conditional statements, such as "if (feature = ...), then it is (ship classification)".

Intelligence Table: Noisy Data for Classification

Name	Planet of origin	Warp Drive Vibration Index (Murds)	Hailing Transponder Freq. (gigaHz)	Surface Reflect. (color)	Ratio of long to short axis	Req. Action
_____	?	7.3	_____	Light Gray	2.1	?
_____	?	6.6	1065.0	White	2.1	?
Lil___	?	6.7	1045.0	White	___	?
_____	?	___	1065.0	Light Color	___	?
Pl__ik	?	7.0	1006.3	Dark Color	___	?
_____	?	7.3	951.4	Green	1.9	?
Krotork	?	7.0	1001.8	Light Gray	1.0	?
7 Woshif	?	___	971.7	Blue	1.7	?
Kritop	?	7.2	____	Dark Gray	2.9	?
C06__	?	6.7	_____	Orange	___	?
0 _____	?	___	____	Black	2.6	?
G__rk	?	6.9	>1000	Black or Dk Blue	3.2	?
_9e__	?	6.6	_____	Light Blue	1.2	?
_6___	?	6.6	_____	Orange	___	?
Rash__	?	___	955.8	Light Blue	___	?
Sor___	?	7.4	<1000	_____	___	?
A____	?	6.8	1013.3	Light Color	1.0	?
E4511	?	___		_____	___	?
_____	?	___	>1000	Light Color	1.7	?
Mor___	?	6.4	1055.0	_____	___	?

For this homework, you will be graded on your implementation(code), write-up, and your classification score.

For your code, make sure the program outputs your classifications.

In your write-up, explain:
- Representation of ship features (f vectors): how you decided which features to use.
- Representation of ship origin (g vectors): what each dimension represents, how you coded them
- Procedure for missing data: your thought process for how to replace missing data
- Learning process: training, testing, and error correcting your system (including all important parameters: e.g., learning constant, convergence criterion, number of trials)

Your write-up should also include:
- clearly stated, the dimensions of your f and g vectors
- a table containing the ship features you used, in the format:

Dimension of f	Name of Feature	Definition of feature	Coding scheme	How you dealt with missing data
1		Written as logical statements (e.g., feature = ???)	binary, bipolar, or standardized values	(e.g., replaced __ with ___)
2				

    *obviously, if there is no missing data for a feature, that cell will remain blank
- your program's output classifications at the end of your write-up

For classification,
getting both the ship origin and required action correct will be considered full score for the classification part of your grade (explain how you did it). Getting the required action correct but the ship origin incorrect receives partial credit. You may use your human pattern recognition skills to classify the new ships and use those inferences to gauge your system's performance. (Getting all ship origins correctly is quite hard and most students couldn't do it, historically. Getting 18 or 19 correct has always been considered as very good.)