Case Study 2
Classification of Cancer Tissues

Download dataset: **http://tinyurl.com/dateset-of-casestudy2**
Alternative link - Click here Data
Note: Before you start Read complete document carefully

Problem Domain

Cancer is severe disease. Early it is detected more chances to save the patient. There are many expert system that are being developing to classify a patient as Cancer patient or not. Accuracy in this kind of problem is very important. As you don't want to make even 0.0001% mistake to say a patient that you don't have cancer. There are many different ways to detect and take observation for same. Finding out the class of tissue can help to make some decision for cancer. Here we have 6 different classes of tissue with 9 observed properties under some experimental situation.

Data Sets

Dataset with electrical impedance measurements in samples of freshly excised tissue from the breast

There are 80 observations and 10 attributes: 9 features+1class attribute

Six classes of freshly excised tissue were studied using electrical impedance measurements:

Code	Tissue name
Car	Carcinoma
Fad	Fibro-adenoma
Mas	Mastopathy
Gla	Glandular
Con	Connective
Adi	Adipose

Impedance measurements were made at the frequencies: 15.625, 31.25, 62.5, 125, 250, 500, 1000 KHz, These measurements plotted in the (real, -imaginary) plane constitute, the impedance spectrum from where the features below are computed.

9 features:

I0	Impedivity (ohm) at zero frequency
PA500	phase angle at 500 KHz
HFS	high-frequency slope of phase angle
DA	impedance distance between spectral ends
AREA	area under spectrum
A/DA	area normalized by DA
MAX IP	maximum of the spectrum

      DR   distance between I0 and real part of the maximum frequency point
      P   length of the spectral curve

Two sheets are given
1. Training Set : with all the features and class of tissue
2. Testing Set : with only features not with class of tissue

Your Job for today:
Make rules to classify – 'Rule Bases Expert System'

Your job for today is to not only to analyze the data and showing the behavior of the data, as you did in case study 1. But also you have to make a system based on rules only, which will classify any tissue for given 9 properties or features into one of the six classes. You have to make your rules, based on given ***training data***.

For example after analysis of training data you may have rule like

"If value of 'I0' feature is greater than _____ and 'DA' feature is less than _____ tissue would be of class _____"

This is again just an example, there may be many 'and' and 'or' 'greater than' and 'less than' clauses in your rules.

Though there is no limitation on number of rules your system should have, but number of rules should be as low as possible. Try to make very comprehensive rules to cover most of cases in short.

This kind of system which work with rules is called **'Rule Based Expert System'.**

After all the rules you made, apply your rules on given ***testing data*** where classification of tissues are not given, just features are given, you have to classify them with your system. After classification submit it to your course instructor and ask for accuracy of your expert system.

Submission

When accuracy of your system is satisfactory, submit your **Model with all the rules** and **report** which shows how did you make that rule based on analysis. All should be in one doc file.