

EE520, Bollt,

**Proj 4, Methods of Machine Learning Problems – Clustering and Classification
Methods - due M Dec 2**

1. Classify the breast cancer images samples – these are human labelled images in two classes. For each, compare the success rates by the cross validation procedure.
 - i. By LDA method
 - ii. By provided “home made” SVM code or using built in SVM (see sample code) fitsvm, and choosing an appropriate nonlinear kernel.
 - iii. By ANN.

I suggest for this problem that you try each of raw images (reshaped as vectors) as features, but then again, also project onto standard basis functions such as initial features, such as FFT or wavelets (fft and dwt2 commands), and finally PCA modes.

2. More Classification. I got from the website on sounds:

<https://www.soundsnap.com/tags/didgeridoo>
https://www.soundsnap.com/music_loops/folk_country

Included - Use the sound import routines in matlab to import at least a dozen sounds in each of two classes and then try out the LDA classification method, (recycling some of the codes above), after projecting onto an appropriate basis set (e.g....is the dwt useful in this setting? fft? POD? Discuss, and choose appropriately). Write a small report style discussion of this using appropriate mathematical formulas, language and supporting figures. So choose two different labeled groups of sounds, to your taste. I chose the didgeridoo music versus folk music, but only because I like to say “didgeridoo” since it is a funny word. You choose as you wish.

3. Clustering. You have been provided with a code that adapts the kmeans clustering method in terms of a well stated metric, as summarized in the also attached paper.
 - A. Restate the metric.
 - B. Interpret the results as coherence, as given for the Rossby wave system. This will include reading a little bit about the Rossby wave equation from whatever source you can find.
 - C. Adapt this same code to work on the Double-Gyre system below – but now it is up to you to also, write out this equation in the usual ODE form, AND look it up from whatever source you can and describe the what it is modelling AND in that context interpret coherence in words (use physical world terms too):.

```
function dy = VectorizedDGode(t,z)
```

```

epsilon =0.1; % set epsilon = 0 for autonomous DG
A = .1;
w = 2*pi/10;
n=length(z)/2;
dy = zeros(2*n,1);
x=z(1:n);
y=z(n+1:2*n);
f = epsilon*sin(w*t).*(x.^2) + (1-2*epsilon*sin(w*t)).*x;

dy(1:n) = -A*pi.*sin(pi.*f).*cos(pi.*y);
dy(n+1:2*n) =
A*pi.*cos(pi.*f).*sin(pi.*y).*(epsilon*sin(w*t)*2.*x + (1-
2*epsilon*sin(w*t)));

```

4. Clustering. Open ended. Included is a spread sheet called Immunotherapy.xlsx. Please read and interpret as you can the columns, and then it is up to you to choose interesting and relevant “features” (either the column values directly, or combinations/functions of them by a scheme you might read about or intuit) and with those, perform a cluster analysis and see if these patients split into distinct clusters. Report what you find. Describe the method(s) you use to do this, and your interpretations of results, in words and graphics.