Homework 6 - Advanced Object Classification

When preparing your homework submissions, don't forget to `git fetch, git status, git pull` *before you issue any other commands in Git – this is to guard against you changing a document that someone else is working on in the same directory*[1].

Don't forget to `git add` *and* `git commit` *(with* `-m` *comments) frequently as you work. This allows other users to see how your work progressed and it automatically backs up your work as you produce. Thus, you're less likely to lose any of your work and/or so you can revert to earlier versions of your work as needed.*

Remember to comment your code carefully with your initials before every comment (as in # JCR `I just wrote a Python comment to document a change.` *Remember to provide an informative header for* **every** *function that you write. Also provide a README file to inform people how to run your code.*

Homework

Objects in the sky can be classified using a number of approaches. Many galaxies are resolved in imaging[2], stars and most quasars are not. Some stars exhibit measurable proper motion, whereas quasars and galaxies typically[3] do not. Quasars have mostly distinct colors from stars and galaxies, which can be even more distinct in deeper or coadded imaging. Quasars, variable stars, and events such as supernovae and flares vary over days to years, but galaxies and most stars typically do not.

The point of this homework is to classify astronomical sources. I will select 5 quasars at redshift $1 < z < 2$ from the *SDSS DR14 Quasar Catalog*[4], 5 variable stars from the *SDSS Stripe 82 Variable Source Catalog*[4], 5 non-variable stars from the *SDSS Standard Star Catalog*[4], and 5 supernovae from the *SDSS Supernova Survey*[4]. I'm not telling you in advance which objects I will select, but they will all have imaging in the *Stripe 82 Catalog Archive Server (CAS)*, they will all lie in the Right Ascension range $30^\circ < \alpha < 60^\circ$, and they will all be representative of their class.

Your homework assignment is to classify these 20 sources as follows:

1. Write code that takes as inputs *just* `RA` and `DEC` (where `RA` and `DEC` can be arrays of any number of rows, i.e. 20 rows!) and returns, as an output, a "classification" array that consists of one column and the same number of rows as the input `RA` array. Each row of the output array should contain a number from 0 to 3, where "0" means "supernova", "1" means "quasars", "2" means "variable star", and "3" means "non-varying star" (or "standard star").

   To obtain useful imaging data for classifying sources from the input `RAs` and `DECs`, your code is allowed to do any or all of the following:

   (a) query the SDSS Stripe 82 CAS to return MJDs, positions, fluxes, and/or magnitudes (and errors associated with these quantities) for sources[5]

[1] This shouldn't be a big deal unless we're working collaboratively, but you should get into the habit *now*.
[2] i.e. they appear extended in imaging
[3] proper motions are increasingly measurable for Local Group galaxies
[4] all of these are linked from the syllabus
[5] see my week12 directory in Git for Python code that can be spawned from within Python to query the CAS

(b) use the SDSS and WISE "sweeps" files that we have stored on `tomservo`.

(c) use the coadd catalog that we have stored on `tomservo`[6], which contains sources classified from a coaddition of all the imaging on Stripe 82 (in order to reduce these sources' flux errors compared to any individual observation).

This week you are NOT allowed to commit any data files to Git. Any data files that your code needs in order to run should already exist on `tomservo` and/or should be generated by the code itself. *If you commit data files to Git, you will receive 0 points for Speed this week.*

I have deliberately imposed these restrictions to prevent you from simply storing all of the classification information from the various catalogs and performing a coordinate match using `search_around_sky`!

To reiterate – unlike the previous homework on object classification, *I will not pass your code any information except an array of `RA` and an array of `DEC` that contain coordinates for each of the 20 objects.*

This week's homework will be assessed objectively, except for commenting which will be assessed in the usual way. This week, accuracy will be worth a total of 30 points, and speed and commenting will be worth 10 points each. Accuracy and speed will be assessed according to benchmarks.

- Accuracy: I will apply your algorithm to an input 20-row `RA` array and an input 20-row `DEC` array, which will correspond to the positions of objects drawn from the 4 classes under consideration. Your score will correspond to how many objects you correctly classify based on your output 20-row array.

  Each correctly classified object will be worth 3 points up to a maximum of 30 points (i.e. 10 correct classifications out of the 20 will be sufficient to achieve the top score).

- Speed: Speed will be assessed out of a total of 10 points using the following formula:

$$t = \text{the total number of seconds your code takes to run}$$

$$\text{IF } t < 60 \text{ then set } t = 60 \text{ ... IF } t > 160 \text{ THEN SET } t = 160$$

$$\text{Score} = 16 - (t/10)$$

  Please include a timer at the very start and the very end of your code that prints out the total time that it takes your code to run.

- Commenting: Will be assessed as usual out of 10 points.

[6] `/astro/astr8020/varcats/varcat-ra30-60-minflux_ngood1-mincoadd_flux_ngood2.fits`