Branch: **master** ▾

Find file   Copy path

ESPM_112L / Week_5 / **Week_5_walkthrough.md**

jwestrob Update Week_5_walkthrough.md

7a551c0 on Feb 20

1 contributor

Raw   Blame   History

111 lines (57 sloc)   6.35 KB

Welcome to week 5 of lab!

This week we're going to be refining our bins and doing some additional binning using ESOMs (Emergent Self-Organizing Maps).

This is a really interesting way of visualizing the way automated algorithms interpret and sort metagenomic data. Next week we'll be using automated binners, which utilize other unsupervised machine learning algorithms. (Not ESOMs, though. We'll talk more about the other popular methods next week.)

This lab is going to be pretty short from your perspective, but know that there was a lot of preparation that went into this week's material! I hope it's enjoyable for you.

Section 1: X11 Forwarding

You're going to be doing this week's lab using a GUI process, which requires some special considerations. GUI stands for **G**raphical **U**ser **I**nterface. You might be thinking, "aren't we doing all our work in a shell?" Well, yes, we are. This week we're going to set things up so that you can run a GUI process on our server while having a window on your computer to modify and mess with as you see fit. X11 forwarding is what you're going to need to use in order to accomplish this.

How to set up X11 forwarding depends entirely on which operating system you're using. In our class, we have four options, which I will order from least to most complicated:

- Linux-based operating systems (NOT CHROMEBOOKS):

  - in your terminal, type `ssh -X [YOUR USERNAME]@class.ggkbase.berkeley.edu` . You're done.

- Mac OS X:

  - For whatever reason, X11 forwarding is disabled by default on macs. Don't ask me why. In order to get around this, you have to go and download XQuartz (https://www.xquartz.org/).
  - Go and download XQuartz-2.7.11.dmg, open it, and open the .pkg file you'll find therein. This will install XQuartz on your system.
  - Now log in and log out of your current session. (Go to the top right, click on your username, and click 'login/switch user' or whatever it says.)
  - Now, go to your terminal, type `ssh -X [YOUR USERNAME]@class.ggkbase.berkeley.edu` . Now you're done.

- Windows:

  - Download and install xming: https://sourceforge.net/projects/xming/

○ Edit Putty preferences:

○ Launch your connection as normal. Congratulations.

- Chromebook:

  ○ Use one of the windows PCs in lab. Sorry.

  ○ If you really are determined to figure out how to do it, look here:
    https://superuser.com/questions/1037230/is-there-a-way-to-use-xterm-on-chromebook

Section 2: Preparing files for ESOMs

Now, I've actually run the ESOMs already. (They take a really long time- best to do that before class.) You just need to gather a couple of things.

Subsection 1: Grabbing all of the ESOM files and putting them in your home directory.

You just need to copy the files over. (Only one person per group needs to do this, strictly speaking, though you all can if you like - the files aren't that large. Talk to me during lab if you're doing this.)

```
cp -r /class_data/[YOUR BABY]/esom_files ~
```

Subsection 2: Generating `.cls` files

What you're going to be doing next is to get a file from class.ggkbase.berkeley.edu (in your browser, not on the terminal) that indicates which scaffolds correspond to each bin.

Only one person per group will need to do this.

Go to class.ggkbase.berkeley.edu, log in, and navigate to the project folder for your baby. You'll be downloading a 'scaffold2bin' file, which has information for each of the scaffolds in your project and their corresponding bin ID. Download it from this drop-down menu:

It is important that you don't rename this file for now!! Don't do it!!

Upload this to your home directory on class.ggkbase.berkeley.edu, either by using Putty to SFTP (Windows) or by using a SCP command on your terminal (Linux/Mac) like so:

```
`scp [BABY ID].scaffolds_to_bin.tsv [YOUR
USERID]@class.ggkbase.berkeley.edu:/home/[YOUR USERID]`
```

Where `[BABY ID]` is the name of your baby (you'll see it in the name of the file you download from ggkbase) and `[YOUR USERID]` is, as you probably guessed, your username (student1, ... , studentX).

Once that's downloaded, you're going to need to convert this to a `.cls` file so that the ESOM program can read it in. You're going to do this using a script I wrote, like so:

```
`python3 /class_data/convert_scaf2bin.py [SCAFFOLDS 2 BIN FILE]`
```

which will create a file called `ggkbase_names.cls` in your current directory. Now, move that file to `~/esom_files` .

Good job. Proud of you.

Section 3: Binning!!!!

Okay, now comes the fun part. You're going to actually run the ESOM binning program! This is the GUI process I was talking about earlier. ONE PERSON per group should do the following, in your home directory (`cd ~`):

- Create a symbolic link to the executable file for the ESOMana program:

  ```
  ln -s /opt/bin/bio/ESOM/bin/esomana .
  ```

- Run the program to launch the GUI

  ```
  ./esomana
  ```

Congratulations! Now let's start loading things. Navigate to the top left of your window and select "File -> Load `*` .wts". You'll find all the files in your `/home/[USERID]/esom_files` folder. If you don't have that folder, that means you didn't follow my instructions from earlier. Go look up.

You'll see a colored map with some dots pop up now, like so:
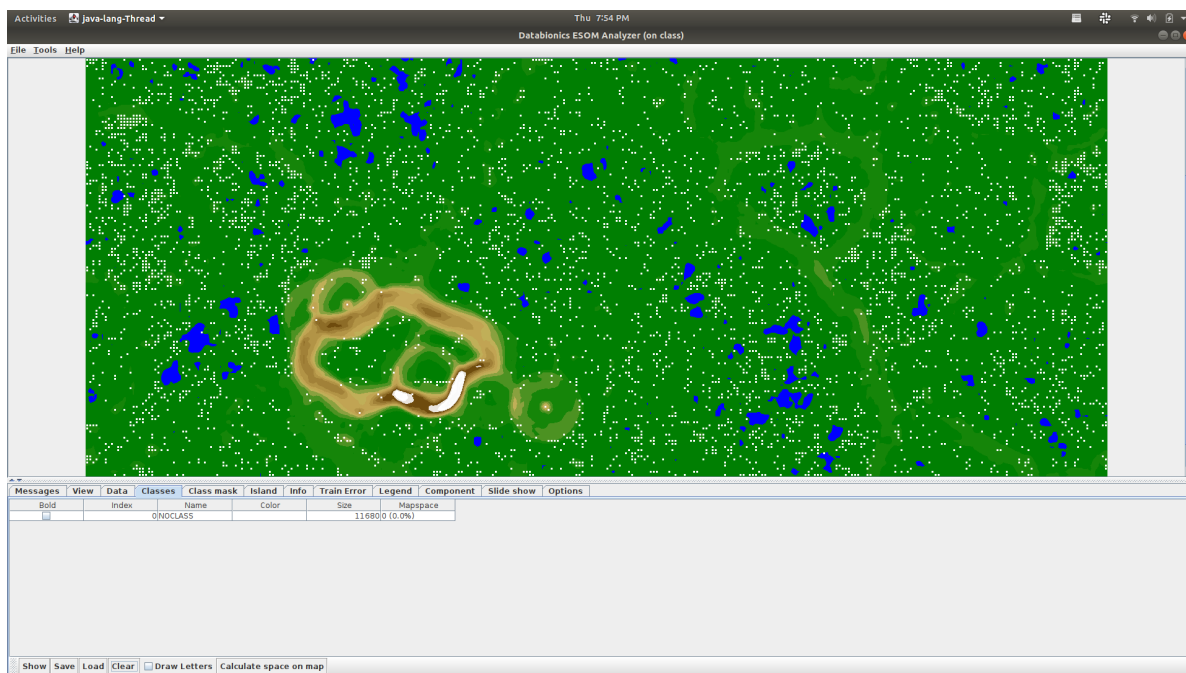
Now you're going to load the class information from your ggkbase bins. Go to "File -> Load .cls" and select "ggkbase_names.cls", which should be in your
`esom_files/` folder.

Now, go and look at the 'Classes' tab in the bottom menu. You'll see a bunch of classes, but everything will have the same color- choose a color for each of your bins. After that's done, you should see something like this:
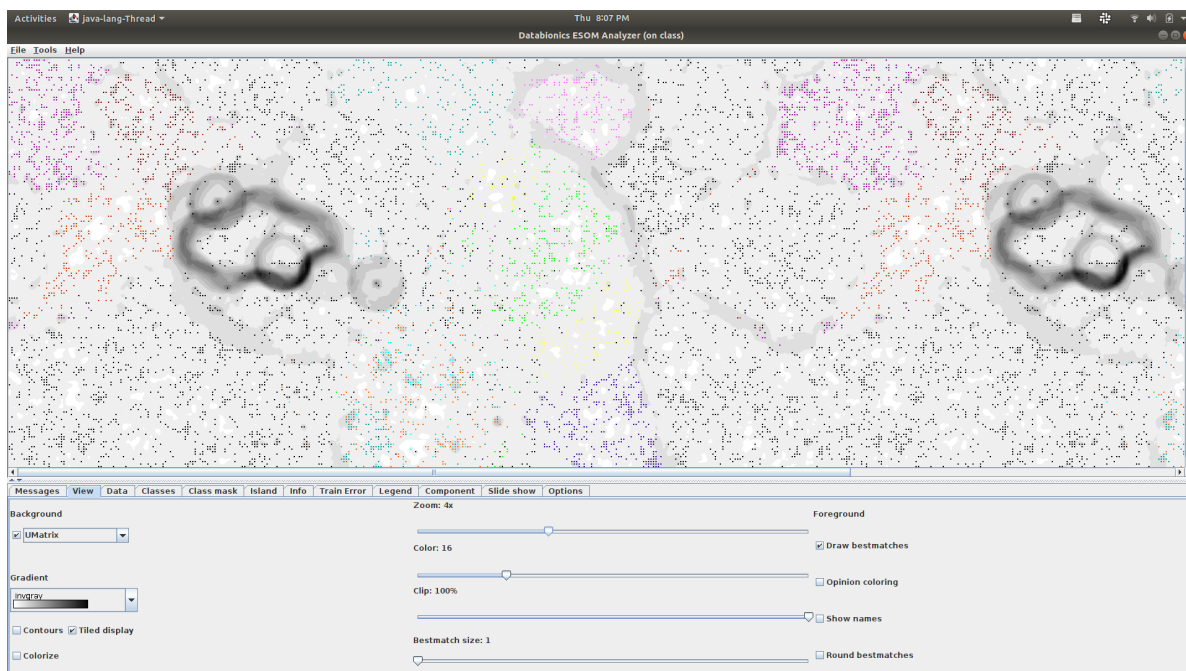
Now we're going to investigate individual bins and see how we can refine them. In your 'Classes' menu, for each class, there's a little box on the left. If you click it, it will make all the points on the ESOM map corresponding to that class bold, for easier viewing. Try it out, perhaps with a couple closely related bins.

Here, I've highlighted two closely related bins (at the family level, so kinda close)... notice how they're next to each other and intermingle a little bit?