GEOG357 Lab 3: Supervised classification

Learn the supervised process and take control in the driver seat.
We'll do it newer cloud-free 2023 PG scene, use the same file as last week.
Submit assignment as .pdf OR .txt

0. Start

If you haven't done it already, update your access to the software by responding to the email requesting you to 'set up your license access' – otherwise you can't start the software.

In Catalyst, Open your copy of the Landsat 5 TM image: **pg20july2023.pix**
Change display to bands 654 and enhance
Expand the image display so it doesn't need panning – but please don't maximise …

Check your pix file for bands/channels:

Change to the 'files' tab and click the + next to rasters; this will show your channels – you should have bands 1-7 and then 5-6 classification/sieve CHANNELS (not bands) from last week.
Switch back to 'maps' tab.

1. Supervised Classification session: Prince George 2023 image

Unsupervised classifications are quick, supervised take longer due to the need to identify cover types and create training sites. I'm picturing 10-15 classes for the image, similar to the class clusters identified last week: e.g. Water, wetlands, Coniferous, Deciduous, Mixed forest, Fields, Cutblocks, Industrial, residential etc.. see the notes below

A. Steps to start supervised classification:

- Analysis -> Image Classification -> Supervised
- Select your 2023 file **.pix ,** click on 'open'
- Click on New Session .. then 'add layer' button
- You need 4 new unsigned 16-bit channels - unless you already have empty channels ?
- Display channels should be 6,5,4 ticked in Red, Green, Blue guns
- Input channels should be 4,5,6; tick in Input channel column
- Set the training channel to an **empty** channel number (tick in that column)
- Set the output channel to another **empty** channel (tick in that column)
- Don't tick training or output in any channel with data already or you will lose it.
- OK

'Training Site Editor' box appears ..

2. Training Areas: *In the Training Site Editor box*

Keep this window open throughout the training session, and move away from the image

We'll start with **Water**, which is usually the easiest to identify:
- Add a new class, (class->new) click on its name and change it to water,
- In the main focus window (directly below Print Icon), pick the icon looks like a zig-zag = 'new shapes', and from the dropdown, pick -> raster seeding.

The raster seeding window controls its functionality with two tolerance settings
These are by default set to 10 and 1x, upping either increases the deviation to include pixels adjacent to wherever you might click.
For water, these might be lower than for other sites, but 10 and 1X are too low. I suggest you try 20 and 4x (for water). I haven't found a big difference between 4 and 8 connect.
Zoom into the image with water and click on the water, I suggest the confluence (T'enneh) of the two rivers, but not too close with the zoom. You want it to find a group of several adjacent pixels, perhaps 5-100.

Select 3-5 seed points for water - you will want to cover clear water (Nechako R) and silty water (Fraser); also pick a lake e.g. Tabor ... you could change training site colours, but DON'T PICK BLUE for water or you can't see your sites - the defaults are usually OK.
The goal is to find a sample of water pixels, not to 'fill' the river.

Q1a: use the help icon (bottom left) to find their suggestion on how many training pixels one should select with 3 input bands. What does it recommend ?
Q1b. What values can the Bias column be- how might you change these to help distinguish water from shadow ?

Throughout seeding, the 'raster seeding panel' MUST be visible onscreen (or you can't seed)
Ensure that the training areas (channel) is highlighted in the table of contents - seeding can be picky. If you lose the raster seeding window, get it back from the zig-zag option
e.g. it disappears if you use the rectangle area zoom in (left of + - buttons)

When done each class**, 'save'**, add the next **new** class, and repeat seeding process by class. You should need to modify the tolerance values for other classes which have more heterogeneity than water, note that the first factor will not go above 50. But read the notes below before you do more, as some classes can be complicated as you discovered last week.

Oooops- what happens when ... you added too much, picked a seed point that spread too much, added something to the wrong class etc.
There are two ways of dealing with this, the easiest is to remove/clear that seed and start over...
- In the training site editing window highlight the class with the mistake
- Select Edit – Clear Selected

You can also use this to remove the entire class e.g. class not intended
- Select Edit – Delete Selected

Usually pick 3-4 seed points for each class training 'set' - and view for different DNs for the same cover type, e.g. due to different illumination (more important in the mountains). For fields and cutblocks, try to pick sites that represent the range of possible values (screen colours). There is overlap between these two classes and also variations due to cutblock regeneration.

As you discovered last week, **coniferous** should class well, being dark and semi-homogenous
Unlike the unsupervised, you should be able to distinguish between residential and industrial.

Fields vs cutblocks is complex without much local knowledge; so I suggest training instead on:
a. Grass = bright red – grass fields / recent cutblocks
b. Pine beetle blocks = brown-green e.g. 5-8km west of campus
c. Mature cutblocks = bright green e.g. one just NE of the beetle blocks on Cranbrook Hill
If you know more about cutblocks and fields, feel free to try and isolate them
How about the power line running N-S between the airport and Tabor Lake ? You could include sections in suitable classes a-c.
You will likely need bigger tolerances for non-water classes – maybe 50 and 16x ?

- Water
- coniferous forest – (include shadowed eskers slopes and dark areas on Tabor Mountain)
- deciduous forest – these show as brighter green e.g. Cranbrook Hill east facing slopes
- mixed forest – can you see 'medium' greens between conif/decid – e.g. on Tabor Mtn ?
- Wetlands (by Swampy Lake, and other similar areas, often at edge of lakes e.g. south of Tabor Lake) – maybe a separate class for the 'damplands' – the bright areas near Swampy Lake
- residential (zoom in to find houses/gardens) – maybe use lower tolerance if it sprawls
- industrial / concrete – look for deep purple areas; maybe include the airport runways
- agricultural fields and recent cutblocks – see notes above for a, b and c
- Parks / recreation: could train on the 15th ave soccer fields, PG golf course, Tabor Mtn ski runs, golf courses north of the Nechako, but they are likely too similar to other green spaces.
- Gravel bars and cutbanks – purple, but brighter than industrial
- chip piles – the brightest features on this image

Any others you can identify ? – remember, you can merge classes afterwards

You could also zoom in on Google maps to help see the landscape in more detail

Very important: when done the sites for each class, be sure to click 'save' then go to class -> new and rename this to a class, then start selecting again – make sure the class you intend, is highlighted. You can also add more areas to previous classes, by being sure to highlight that class and seeding further.

When you're done, in the **training sites window**, select **tools -> signature separabilities**
for suitability of classes and training areas - ideally each matrix value is > 1.8
Expand the window to see all the sig.sep. values
If some pairs are low (below 1.5) you can either delete one or merge the pair if it's logical
Class->merge select source classes and the destination class (one of the source classes), Apply
- this is only a lab exercise, so no need for excessive rigour, but merge two classes if the SS stands out as an anomalous low value. You could then review the revised SS report file.
Save the signature separability report, as a **.txt** file

Q2. comment on any low separability values – and if 'easily' fixed, e.g. did you merge the two, or modify training samples – otherwise comment on why they have happened

Tools-> classification preview
This will show how it will look .. go back if it looks doomed and make edits as needed e.g. new classes; but usually it should be OK for the lab exercise.

You have these options - try them all as they are previews and show immediately: Click off /on the classification layer to view the 654 composite underneath to check how each compares with the land cover types; the PseudoColor table (PCT) is random, so try to see through the colours – you could edit a few if it helps, e.g. blue for water, dark green for coniferous etc..

- Maximum likelihood
- Maximum likelihood with null class (null = black) – what is missing, do we need more ?
- Parallelepiped
- Parallelepiped with MLC tiebreaker
- Minimum distance

Do they retain the airport runways, do they avoid classing the esker shadows as water, how well do they separate residential from industrial, do they identify the gravel bars along the river ?

Remember that urban classification is challenging due to the mosaic of many small features – and that some features look similar e.g. agricultural and sports fields, and fields / cutblocks.

3: Run the Classification

Pick the best classification based on observation of areas you know – I was watching for the eskers area – did one or more avoid confusing the esker shadows with water ?
This is all a bit subjective without having ground truth information
As per lecture notes, one might expect the best to be the MLC option .. but then maybe not ?

right-click on the **classification metalayer** in the 'maps' list and select **Run Classification**.
Select your preferred classification. Make sure the show report button is ticked. Note the 'use bitmap as mask' option – we won't use it here to restrict a classification to only part of the image.
I**n the report window** which classes are most confused (in the confusion matrix) ?
- why are these classes confused and not others? (maybe none!). Note that the accuracy shown is inflated as this is only comparing with training areas, which should mostly be correct.
The **confusion matrix** gives a rough guide on the classification based on the training sites
Save this report as a **.txt** file … retain only down to the line with Kappa value. This enables you to refer to it if need be after the report window is closed.
Reference data listed in the columns of the matrix represent the number of correctly classified training samples. Errors of omission are represented by the non-diagonal **column** elements, and errors of commission are represented by non diagonal **row** elements.
The Accuracy Statistics report lists statistical measures of overall accuracy and per class.

Q3a. what is your kappa value ?
3b. which class has the most pixels and which has the fewest ?
3c. comment briefly on the least and most confused classes with possible explanation why

Now run a second classification type into the next empty channel.

Right-click on the **classification metalayer** and select **configuration –** change the output channel to the next empty channel, so it doesn't overwrite the previous one.
Then run the classification again using your second preferred method.
again save the report text file (down to the Kappa value)

Q4a What is your Kappa value
4b. Now which class has the most pixels and which has the least – is it the same as in Q3 ?
4c. briefly comment on how this compares with your first classification – e.g. do either (or both) help distinguish the dark slope shadows versus water ?

You can display the previous classification if you need to using: Layer- Add -> pseudocolor
Then select that classification channel;
you could also load your unsupervised classification from lab 2 the same way for comparison.

4. SIEVE

SIEVE parameters :
input = ## (your preferred classification channel) .. note –NOT the training channel
Polygon size threshold = 5 (for 1 acre minimum), or 11 for a hectare...
Connectedness - can be 4 or 8
exclude values list = ## (where ## is the class number for water – to preserve small lakes)
output port should be viewer - PCT
select log tab and run ...

View the result, compare with the unsieved classification (tick the sieved layer on and off)
If happy with the result, Re-run the sieve putting output to your pix file – use BROWSE button - remember everything goes into your 2025pix file, don't put it somewhere else … Run
This will create a new SIEVE channel

View your raster layers (files tab) listing to check you have a sieve layer – It may be labelled 'contents not specified' .. if so, go to the files tab, double-click on the name and change to something like 'sieve to 1 acre'. [if you rename under the 'maps' tab this only changes the name onscreen- the channel remains unnamed if you close it]
Under no circumstances, should you have a new file named sieve.pix – a useless orphan file 😊

Compare with your unsupervised classification:
Layer->add->pseudocolor- and select your 'best' unsupervised channel from last week –> finish

It's not easy to compare as the class colours don't match, but you should find more 'order' in today's result as you had more control over the classes. You are the boss !

Q5. Briefly describe two situations where your supervised classification seemed to perform better than the unsupervised – it could be separating certain classes for example. Comment also on any situations where the unsupervised seemed better e.g. in finding unexpected clusters.

This webpage might help in directing your thoughts:
https://mapasyst.extension.org/whats-the-difference-between-a-supervised-and-unsupervised-image-classification

Please submit Lab 3 assignment as .text or .pdf by Friday 5pm

this section is for reference only and possible use in projects.

Appendix. Full Accuracy Assessment (time consuming and picky-tricky)

How could you more fully assess your classification - you'd need some independently derived 'ground truth' plots to test against: (these would usually be generated from ground work or other knowledge). The purpose of the accuracy assessment is to create a report which indicates the accuracy of classification results compared to the raw image data. It compares what is assumed to be correct with an image classification based on pixel groupings.

Select Analysis dropdown -> Image classification -> Supervised-> Post Classification Analysis-> select "Accuracy Assessment".
Select your file (yet again)
In the new accuracy assessment window, click on 'select classified image'
And choose the **Sieved** results of the Supervised Classification
Select "Load Reference Image" and load channels 5, 4, 3. … OK
This image will be used during the random sample selection.

Generating Random Samples

Select "Generate Random Sample" and increase the sample number to 30.
Leave the "Sample Control Options" section to the default answer, "yes". Press "Accept".
In the new Accuracy Assessment Panel, the idea is to highlight the first point in the sample list and observe where it is located on the image (the cursor will automatically be placed at this location). You should likely zoom in to both find the cursor and identify the class type (as best you can). Compare this pixel to the list and select the class you believe it should belong.

Ideally the pixel is in a relatively homogenous (not mixed) area .. but if this is not the case, you can ignore/reject that pixel. Once the random sample you are working with has been highlighted, you can simply select the class it belongs to (click on that class) and the information automatically transfers to the random sample list. This works the same as the "Transfer" button.
Samples could also be taken from vector segments e.g. verified vegetation polygons.
The user is given the option of selecting the channel and the class attribute associated with it.

Produce Accuracy Report
Once all 30 samples have been assigned, select "Accuracy Report" then 'generate report'
The results of the accuracy assessment are shown in the form of a Sample report Listing, Error (confusion) matrix and as accuracy statistics. Don't be discouraged by low values – in a project/thesis you would take many more samples to get higher stats – at least 30 per class.
The Sample Report Listing can be used to determine which samples were classified correctly.
You would save this report as a .txt file in your folder for later reference.