FYS3023: Environmental monitoring from satellite, Fall 2018

Assignment 2:
Mini project

Xu Xu
xu.xu@uit.no

Andrea Marinoni
andrea.marinoni@uit.no

Deadline for hand in: 14:00, 30 October (Tuesday) 2018

Before we start...

...what to submit?

- Without surprise, your task is to implement classification with SAR data. The first problem is about supervised classification method, and the second one is an application of unsupervised classification method. The exercise should be presented by uploading a PDF document on Wiseflow before the deadline (14:00, 30 October (Tuesday) 2018). The document must be *written individually*, but you are free to discuss the problems together.

- There will be a short (around 10 mintues) oral presentation to the fellow students about what you have done about this part. You need to do both the problems, but you only need to choose one problem for presentation. This is mandatory but will not be graded, the main point is to learn from each other! (For now the scheduled time for presentation is 5 November on the lecture class.)

- Data can be found on Canvas or using the link http://t.cn/Ehg4Wo4.

The grading will be based on your discussions and comments. Therefore, make sure to be clear and precise when discussions and comments are asked for!
Code, an important part, should be attached as an appendix, make sure it is well commented and easy to follow. We will read your code to check if you do it individually.

Good luck!

1 Mr X's potatoes

Mr X is a potato farmer in the Flevoland area in the Netherlands. He is from a family with a long history in the potato business. Mr X is very much up to date with the latest advances in remote sensing. Recently, he has been very interested in what Synthetic Aperture Radars (SAR) can achieve, and realised that this is a tool that can help him monitoring his potatoes. He has contacted the European Space Agency, which in turn has assigned *you* to help him.

Your task is to implement a supervised land cover classification algorithm for polarimetric SAR data. You are given a test image over Mr X's potato fields, accompanied with a ground truth map.

Do the following:

1 a) Load the fully polarimetric SAR image stored in "flevoland.mat". The file contains the elements of a multi-looked covariance matrix $\langle \mathbf{C} \rangle$ and a ground truth map specifying the following land cover types:

0 = Unknown	8 = Bare soil
1 = Stem beans	9 = Beets
2 = Forest	10 = Rapeseed
3 = Potatoes	11 = Peas
4 = Lucerne	12 = Grass
5 = Winter wheat I	13 = Water
6 = Winter wheat II	14 = Summer barley
7 = Winter wheat III	15 = Buildings

The data for covariance matrix C is a 3*3*745*1024 matrix. So C(:,:,1,1) is the covariance matrix for the first pixel if you look at the data in Matlab. Another two variables are for the ground truth: a class map and a colour table. You can use imagesc and colormap to show it. Please show the RGB image of the data using the diagonal elements of the covariance matrix, and the ground truth. (1p)

1 b) For each set of pixels with a known ground truth label, i.e. for each ground cover class, split the pixels into two sets randomly: one set used for training your classifier (called the *training set*) and the other set used for assessing accuracy (called the *test set*). For example, use 60 percent of pixels as training set and 40

percent of pixels as test test for each class. (1p)

1 c) We shall assume that the intensity distributions for each class can be modelled by a gamma distribution, which is the most commonly used probability density function (pdf) model for multi-look intensity data covering a homogeneous area. So, a multi-looked scattering intensity (i.e. the diagonal elements of $\langle \mathbf{C} \rangle$), call it I, can be modelled statistically by the Gamma distribution:

$$p(I; \alpha, \mu) = \left(\frac{\alpha}{\mu}\right)^{\alpha} \frac{I^{\alpha-1}}{(\alpha-1)!} \exp\left(-\frac{\alpha}{\mu} I\right) \tag{1}$$

where the two parameters α and μ are model parameters. In general, both model parameters are class dependent, but for this given image you may treat α as a class independent constant. Specifically, use $\alpha = 27$ for all classes. About μ, it can be estimated using the maximum likelihood. It can be easily derived that μ is the average of the intensity data.
Use the training data to estimate μ and the Gamma function for each class. To check if Gamma function can represent the distribution of the training data for each class, plot the histogram of the intensity values (The histogram needs to be normalized to have a total integral as 1). On top of it, plot the Gamma distribution. Since we have a lot of classes, please just choose class 3(potato), class 5(winter wheat 1) and class 15(building) to do the check. Do the estimated Gamma function fit the data? Make some comments. (3p)

1 d) Now we can use the Gamma function and the training set to train a *maximum likelihood classifier*, considering all known ground cover classes. The maximum likelihood classification is based on the Bayesian decision theory, so we need to build a discriminant function. Please show the expression of the discriminant function (This function should be simplified and expressed without α). Use the discriminant function to do the classification and show the classification results. (5p)

1 e) What is the accuracy for Mr'X potato? Use the test set to assess the accuracy. Please consider the confusion matrix and discuss possible errors. Notice that there are actually 3 intensity data (HH, HV, VH) that we can use, which channel has the best classification result?(4p)

2 San Francisco needs your help

Mr Y is a friend of Mr X. Mr Y is working for the land management department at San Francisco. Mr X recommends you to Mr Y to help with the classification of San Francisco area. Obviously Mr Y is a bit smart. He notices that the method for Mr X's problem only uses the information of intensity and doesn't take full advantage of polarimetric information. So for this task, you need to generate some polarimetric features and design an unsupervised classifier.
The dataset is the sanfransisco.mat. The scattering matrix of a SAR image is stored in this file. The scene shows the City of San Fransisco, USA. From the google map of this area, we roughly divide it into 3 classes: open water, urban area and forest (e.g. the Golden Gate Park).

Do the following:

2 a) Show the RGB image using the diagonal elements on the covariance matrix and coherency matrix (You can directly use your codes for the last problem in assignment 1).

2 b) Firstly, we generate a polarimetric feature that can divide the image into urban area and other area. Create a ROI (Region Of Interest) for urban area (draw a polypon on urban area, use roipoly in matlab). Considering all the pixels within the ROI, compute the first principal component of the target vector (k_l in assignment 1) by using the Principal Component Analysis (PCA), call the first principal component $\hat{\mathbf{c}}_1$ (you may use built in functions, like *pca*, *pcacov* or *eig* in Matlab). Make sure the component is of unit length.
For each pixel in the full scene, compute the angle θ between the target vector of the pixel, k_l, and the principle component $\hat{\mathbf{c}}_1$, as:

$$\theta = \cos^{-1}\left(\langle \frac{|k_l^\dagger \hat{\mathbf{c}}_1|}{|k_l|} \rangle\right)$$

where and † means transpose the ⟨⟩ means average. You can use a sliding window to achieve it like multi-look. Display the angle as an image, where the contrast is stretched between 0 and $\pi/2$. Explain why some regions yield high values and some low values. (2p)

This is a feature derived from the target vector and can be used to detect the urban area. More features with physical meaning (e.g. related to scattering mechanism) can be extracted by decomposing the coherency matrix.

Do the following:

2 c) Compute two features from the coherency matrix: the alpha angle $\bar{\alpha}$ and Entropy H. Make a scatter plot in the $H/\bar{\alpha}$-plane and try to find thresholds on H and $\bar{\alpha}$ which clusters the SAR data into 3 classes: urban area, forest and open water. Do the classification based on the thresholds you choose and show the classification result. (Hint: when you do the scatter plot, randomly choose a portion of pixels for plotting to reduce computational memory.)(4p)

2 d) Design an unsupervised method to do the classification. You can choose one the classic methods (e.g. Fuzzy C-means clustering, Wishart...). The polarimetric features for use could be the alpha angle $\bar{\alpha}$ and entropy H. Or you can use some other features (e.g. anisotropy...). You should also try to divide the data into more classes (more than 3) and see if your classifier can help Mr Y get more information. Compare your classification results with google map and comment on your results. (5p)