Soft Computing:ISCO630E

Report

Assignment 5.2

Anant Chaturvedi(IIT2016506)

Question Description

2. Using Naive Bayesian classifier predict river non river using Satellite data set of Hooghly river (unstructured data set).

Concepts Used:

GAUSSIAN CLASS CONDITIONAL DENSITIES

Here, $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \{1, \ldots, K\}$. Estimate Bayes classifier via MLE:

- **Class priors**: The MLE estimate of π_y is $\hat{\pi}_y = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(y_i = y)$.
- **Class conditional density**: Choose $p(x|Y = y) = N(x|\mu_y, \Sigma_y)$. The MLE estimate of (μ_y, Σ_y) is

$$\hat{\mu}_y = \frac{1}{n_y}\sum_{i=1}^{n} \mathbb{1}(y_i = y)x_i,$$

$$\hat{\Sigma}_y = \frac{1}{n_y}\sum_{i=1}^{n} \mathbb{1}(y_i = y)(x_i - \hat{\mu}_y)(x_i - \hat{\mu}_y)^T.$$

This is just the empirical mean and covariance of class y.

- **Plug-in classifier**:

$$\hat{f}(x) = \arg\max_{y \in \mathcal{Y}} \; \hat{\pi}_y |\hat{\Sigma}_y|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \hat{\mu}_y)^T \hat{\Sigma}_y^{-1}(x - \hat{\mu}_y)\right\}.$$

Taken from the slides of the edX course ColumbiaX: CSMM.102x Machine Learning

Introduction

We are given 4 images denoting the R band,G Band,B Band and I Band Images of an image with river and non river points.We have to classify every point in a 512X512 image into either a river class or a non river class.

Steps Followed:

•Four satellite Images of Kolkata (Rband, Gband, Bband and Iband) are given to

you with equal image size (512 * 512).

• The feature vector dimension is 4

• Each pixel location we have four values.

• Two Classes are given (River and NonRiver)

• Take 50 sample points (Pixel location's corresponding pixel values) from river class for training for each band

•Take 100 sample points (Pixel location's corresponding pixel values) from non

river class for training for each band.

•Take (512 * 512) sample points (Pixel location's corresponding pixel values) for

testing for each band.

•Apply baye's decision rule to classify all the test sample either in river or nonriver

class denoting 0 and 255 at corresponding pixel locations.

•Show the result in image form with black and white image (either 0 and 255)

•Step 1: Calculate Mean of River Class : T1 = [Mean1; Mean2; Mean3 ;Mean4];

Mean1 = mean of Rband image for 50 sample points

Mean2 = mean of Gband Image for 50 sample points

Mean3 = mean of Bband image for 50 sample points

Mean4 = mean of Iband image for 50 sample points

•Step 2: Calculate Mean of NonRiver Class : T2 = [Mean1; Mean2; Mean3; Mean4];

Mean1 = mean of Rband image for 100 sample points

Mean2 = mean of Gband Image for 100 sample points

Mean3 = mean of Bband image for 100 sample points

Mean4 = mean of Iband image for 100 sample points

•Step 3: Calculate the Covariance Matrix for River Class for 50 samples which is 4 * 4 dimensions. Basically (X – T1) deviation and (Y – T1) deviation and multiply it and summing up where X and Y represents all the sample points considered for training (R, G, B and I band image) we will get 2^4 = 16 values in the covariance matrix for possible combinations of 4 band images. We are doing the deviation of sample points from the mean vector.

(Apply covariance matrix calculation formula)• Step 4: Calculate the Covariance Matrix for Non River Class for 100 samples which is 4 * 4

dimensions also by applying same process explained in step 3.

• Step 5: Take whole image for test data where : test_data= [Rband_img(i,j) Gband_img(i,j) Bband_img(i,j) Iband_img(i,j)]; i = 1 to 512; and j = 1 to 512;

• step 6: The dimension of test data is (4 * (512 * 512));

• Step 7: For each pixel location of test image Run the loop from i = 1 to (512*512) Do

• Step 8: For river class calculate (test_data – T1) deviation and (test_data – T1) T Then Multiply it :

River_class = (Test_data – T1) T * Inverse (Covariance_matrix_Riverclass) *(Test_data – T1)

• Step 9: For Non_river class calculate (test_data – T2) deviation and (test_data – T2) T Then Multiply it :

Nonriver class = (Test_data – T2) T * Inverse (Covariance_matrix_NonRiverclass) *(Test_data – T2)• Step 10: Calculate density function p1 for river class where P1 = 0.3 given

p1 = (-0.5) * 1/sqrt(Determinant of Covariance_matrix_Riverclass) * exp(River_class);

(Here we apply multivariate Normal Distrubution)

• Step 11: Calculate density function p2 for nonriver class where P2 = 0.7 given

p2 = (-0.5) * 1/sqrt(Determinant of Covariance_matrix_nonRiverclass) * exp(NonRiver_class);

• Step 12: For each pixel location of test image apply baye's rule (P1 * p1) >= (P2 * p2) then

Out_image(i) = 255 (River class)

Else

Out_image(i) = 0; (Nonriver class)

• Step 13 : Goto step 7;

• Step 14: Show the three output image Image using imshow function for three cases:

Case 1 : River class (Prior Prob:) = 0.3 , Nonriver class(Prior Prob) = 0.7

Case 2 : River class (Prior Prob:) = 0.7 , Nonriver class(Prior Prob) = 0.3

Case 3 : River class (Prior Prob:) = 0.5 , Nonriver class(Prior Prob) = 0.5

RESULTS

For P1=0.7 and P2=0.3 the results are :-