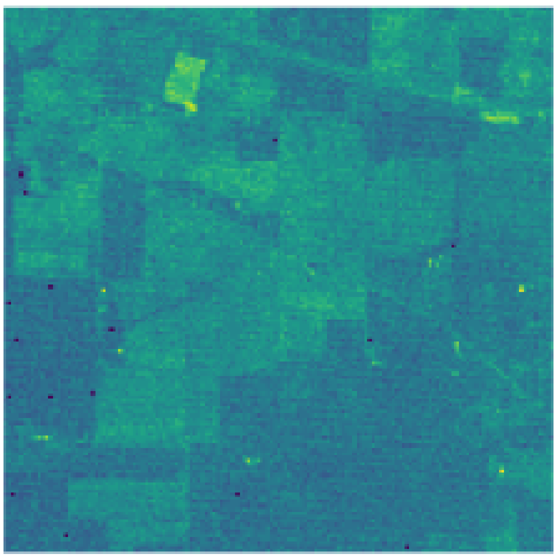
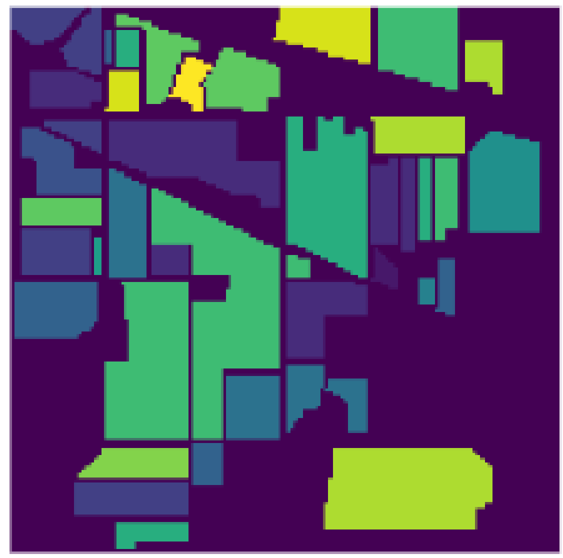
GNR 602

Image Segmentation using Fisher's Linear Discriminant Classifier

Naman Singh Rana, Utkarsh Ranjan and Utkarsh Pratap Singh



Dataset



Ground Truth

Introduction

We used two algorithms , Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), to perform image segmentation on satellite images. Among numerous feature extraction and classification methods, Principal Component Analysis and Fisher's Linear Discriminant Analysis are the most popular, relatively effective and simple methods and widely used . It is generally believed that, when it comes to solving problems of pattern classification, LDA-based algorithms outperforms PCA-based ones, since the former optimizes the low-dimensional representation of the objects with focus on the most discriminant feature extraction while the latter achieves simply object reconstruction.

A distinct difference can be made between an unsupervised statistical method such as PCA and supervised methods such as LDA. While PCA is useful for finding components that represent data, there indicates no evidence that these components are useful for discriminating between different classes. There is a chance that the bands or directions discarded by PCA are the ones that can discriminate between classes. LDA is an approach that can maximize between-class separation while minimizing within class scatter .

Dataset

We used Indian Pines Dataset, a remote sensed image for land use land cover. This was gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of 145\times145 pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5 10^(-6) meters. This scene is a subset of a larger one. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. There are two major dual lane highways, a rail line, as well as some low density housing, other built structures, and smaller roads. Since the scene is taken in June some of the crops present, corn, soybeans, are in early stages of growth with less than 5% coverage. The ground truth available is designated into sixteen classes and is not all mutually exclusive.

Methodology

Hyperspectral image (HSI) contains rich information from a broad electromagnetic spectrum that can be used to identify or classify target objects with subtle differences. Due to the higher dimensional nature of HSI, the classification system often requires a feature extraction/selection process followed by a supervised learning classifier. We studied Principal component analysis (PCA), Fisher's linear discriminant analysis (LDA) based on distance measures for Preprocessing with different classifiers such as K-means, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) for Training. In Postprocessing, to Smoothen out the final prediction matrix we used Gaussian Blur for noise reduction. We have 200 data channels and with LDA and PCA we reduced dimensions to 16 using PCA and LDA.

Fisher's Linear Discriminant Analysis

Assuming an HyperSpectral Image that has m number of training pixels represented by $\{p_1, p_2, \ldots, p_m\}$ for N number of classes, i.e., $\{n_1, n_2, \ldots, n_N\}$, and there are m_j number of samples in the j th class, i.e., $\sum_{j=1}^{N} m_j = m$, let μ be the mean of the training set and μ_j be the mean of the j th class. Then, the within-class (S_W) and between-class (S_W) scatter matrices are defined in equations 1 and 2.

$$S_W = \sum_{p_i \in n_j} (p_i - \mu_j)(p_i - \mu_j)^T \qquad (1)$$

$$S_B = \sum_{j=1}^{N} m_j (\mu_j - \mu)(\mu_j - \mu)^T \qquad (2)$$

The goal of LDA is to find a transform w that maximizes the Raleigh quotient (Q) given in equation 3.

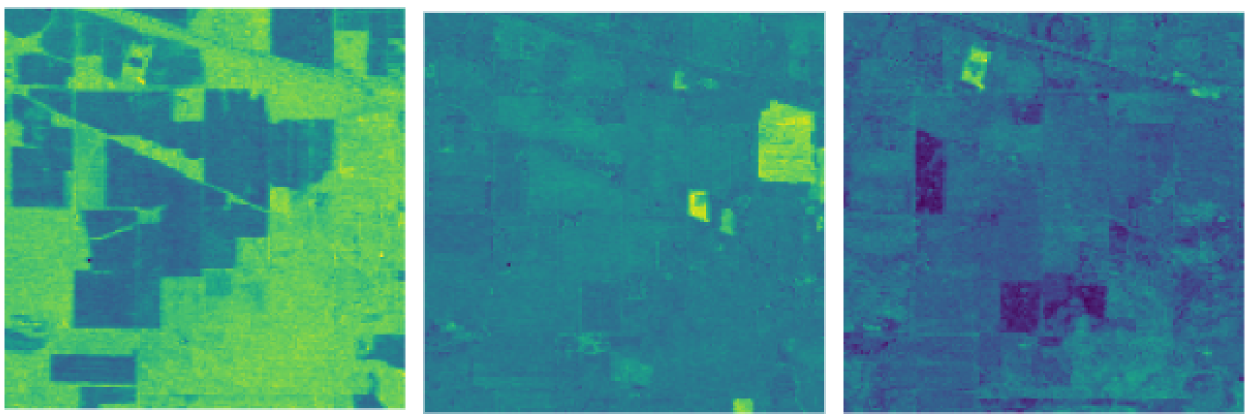
$$Q = \frac{w^T S_B w}{w^T S_W w} \qquad (3)$$

The transformation matrix w can be determined by solving the generalized eigen-problem given in equation 4.

$$S_B w = \lambda S_W w \qquad (4)$$

Where λ is a generalized eigenvalue and W is the transformation matrix that reduces the dimensionality of HSI data from D to N - 1 with maximum class separation in the lower dimensional space.

LDA(dominant 3 components)



Principal Component Analysis

We first preprocess the data by normalizing each feature to have mean 0 and variance 1. We do this by subtracting the mean and dividing by the empirical standard deviation:

$$x_j^{(i)} \leftarrow \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

Where $\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_j^{(i)}$ and $\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n} (x_j^{(i)} - \mu_j)^2$ are the mean and variance of feature j, respectively.

To maximize the variance of the projections, we choose a unit-length u so as to maximize:

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_j^{(i)T}u\right)^2$$

Maximizing this subject to $||u|| = 1$ gives the principal eigenvector of

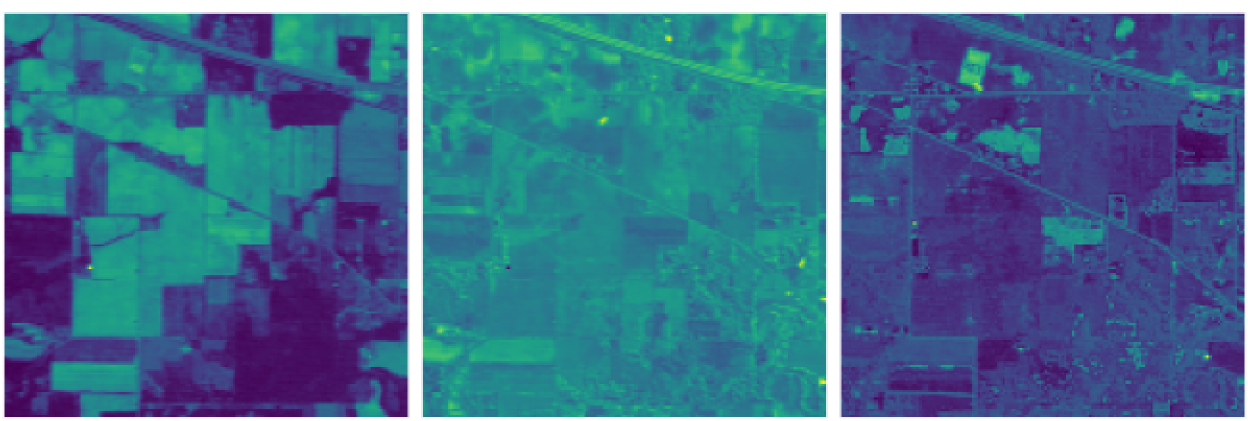
$\Sigma = \frac{1}{n}\sum_{i=1}^{n} x_j^{(i)} x_j^{(i)T}$, which is just the empirical covariance matrix of the data (assuming it has zero mean).

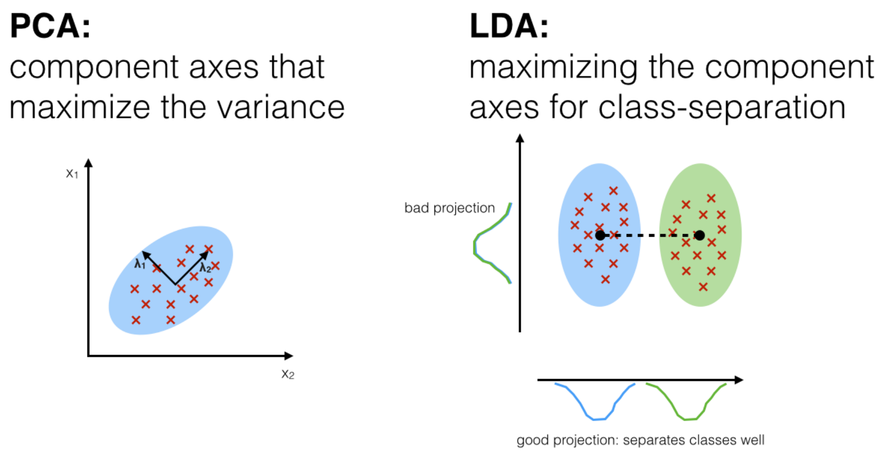
Then, to represent $x^{(i)}$ in this basis, we need only compute the corresponding vector $y^{(i)}$.

The vectors $u_1, \ldots, u_k$ are called the first k principal components of the data.

$$y^{(i)} = \begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \vdots \\ u_k^T x^{(i)} \end{bmatrix} \in \mathbb{R}^k.$$

PCA (dominant 3 components)

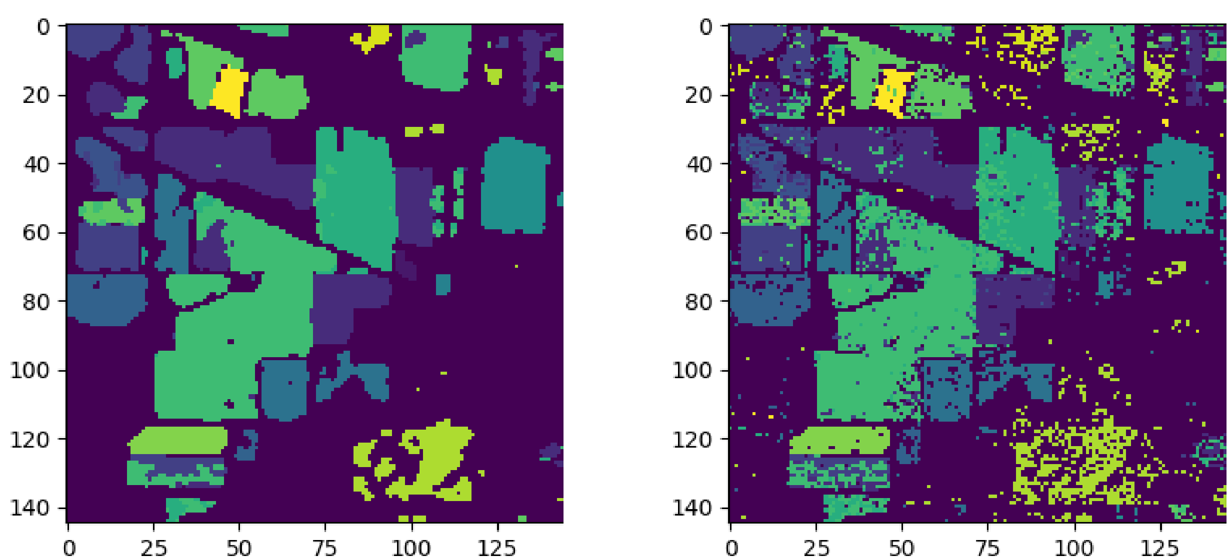




*

Gaussian Blur

The Gausssian filter is a type of linear filter that uses a Gaussian function to compute the weighted average of neighboring pixel values. This function assigns higher weights to pixels closer to the center of the filter and lower weights to those farther away.When applied to an image, the Gaussian filter blurs the edges and reduces the high-frequency content, effectively removing noise and enhancing low-frequency details. Left image is after application of Gaussian blur.



PCA vs LDA

PCA is a Unsupervised learning technique for dimensionality reduction. It finds a lower-dimensional representation of the data that retains as much of the original variation as possible.But it does not take into account class labels.

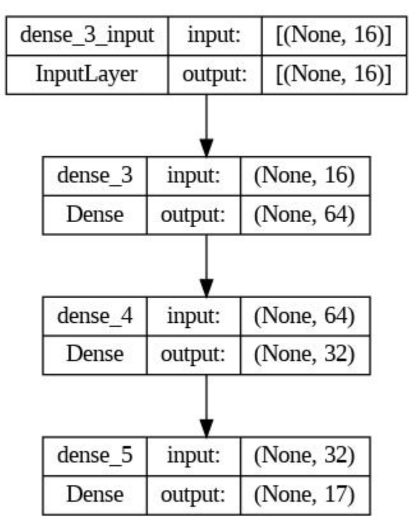
LDA is a Supervised learning technique for dimensionality reduction. It finds a lower-dimensional representation of the data that maximizes separation between different classes. Thus it takes account of class labels.

The performance of LDA and PCA based classifiers depends on the dataset and the classification problem. In general LDA outperforms PCA when there is a clear separation between the classes. On the other hand, LDA can be more effective when there is a significant overlap between the classes.

Training

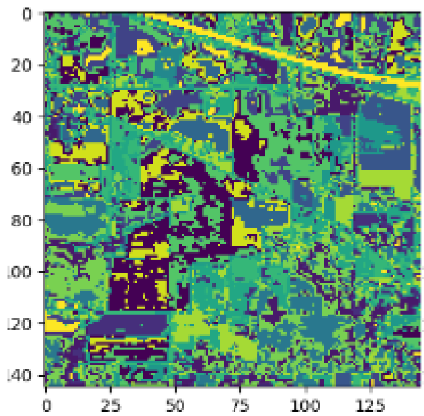
We split the data randomly with 70 % pixel used in the training set and the rest 30% for testing. We used Support Vector Machines and Artificial Neural Network as classifiers. And finally used Gaussian Blur to smoothen the prediction matrix obtained from Neural Network.

Models

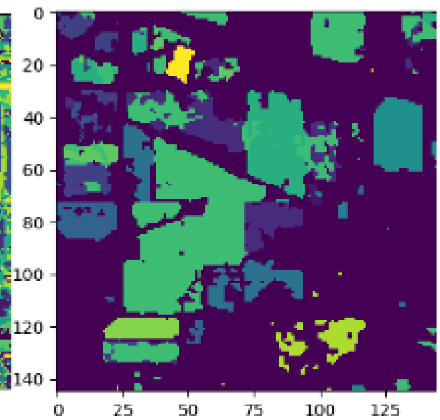


Outputs

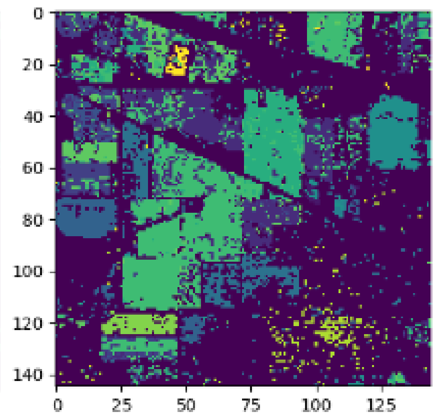
Outputs (PCA)



SVM

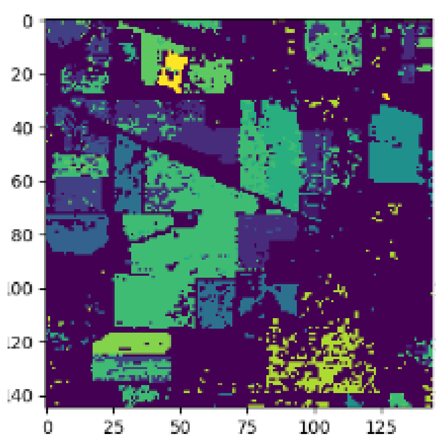


ANN + Gaussian Smoother

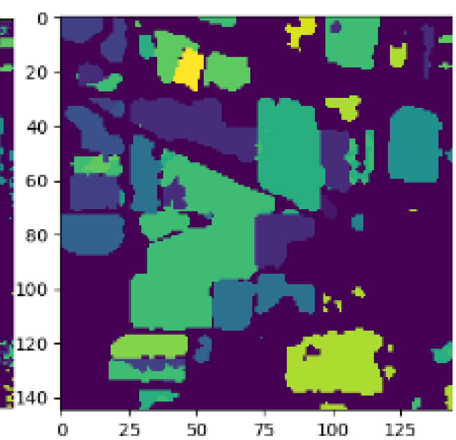


ANN

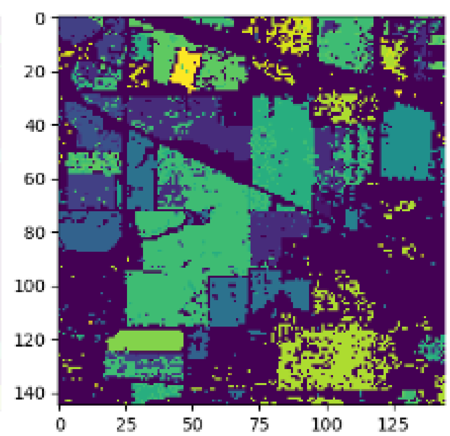
Outputs (LDA)



SVM



ANN + Gaussian Smoother



ANN

Results (Indian Pines)

Accuracy

Model	PCA	LDA
SVM	64.16%	74.67%
ANN	70.12%	77.24%
ANN + Gauss Smoothening	78.12%	**85.21**%

MSE

Model	PCA	LDA
SVM	32.553	23.3199
ANN	28.419	17.4061
ANN+ Gauss Smoothening	28.595	17.0469

Results (Salinas)

Model	PCA	LDA
SVM	73.84%	83.16%
ANN	82.26%	89.69%
ANN + Gauss Smoothening	83.95%	**93.15**%

Observations

Best prediction results obtained using **ANN + Gaussian Blur** with **LDA** (85.21%)

LDA reduces the dataset while preserving the **class-discriminative** information, whereas PCA only captures the directions of **maximum variance** in the data. This makes LDA **better** suited for classification tasks and can lead to better performance compared to PCA.