Gender Classification using LDA, KNN, SVM, Naive Bayes and Decision Tree Learning

SUBMITTED BY

NAME: MAYANK WADHAWAN

UFID: 59148122

Topics Covered

- Applications
- Introduction
- Data Set
- Classifications Used
- Running Code
- Existing Approaches
- Results
- Future Improvements
- Conclusion
- Project Demonstration

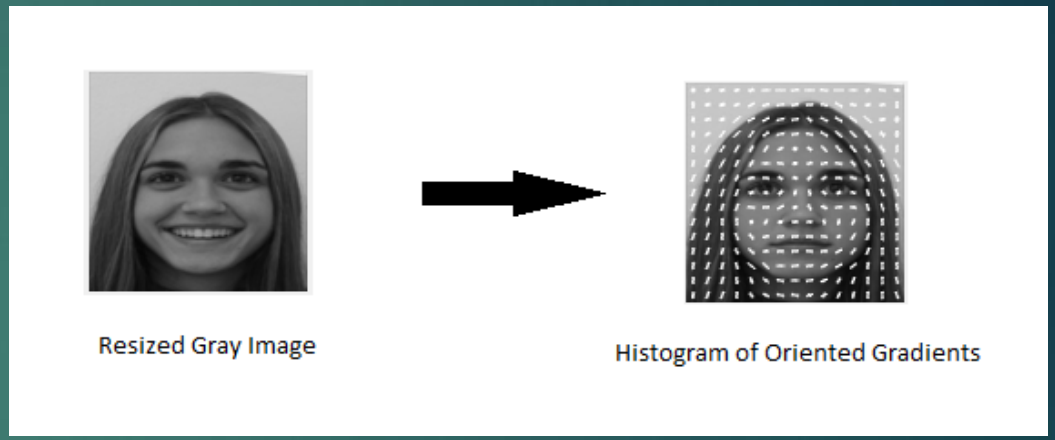
Applications

Gender classification has many applications like:-

- Targeted advertisement
- Video Games (Avatar)
- Statistics

Introduction

- Gender classification on images.
- Firstly, there is learning phase
  1. 312 images read from folder
  2. Converted to gray
  3. Gaussian filter applied
  4. HOG features extracted
  5. Data fed into 5 different classifiers
     - LDA, KNN, SVM, Naïve Bayes and Decision Tree Learning



Introduction (continued)…

- Secondly, Prediction is done
  1. Test image read from gallery
  2. Converted to Gray
  3. Gaussian Filter Applied
  4. HOG features extracted
  5. Sent to classifiers for prediction
  6. Result verified by comparing actual label with predicted label

Data Set – FEI Face Database

	Male	Female
Training Images	156	156
Testing Images	40	40
Total Images	196	196

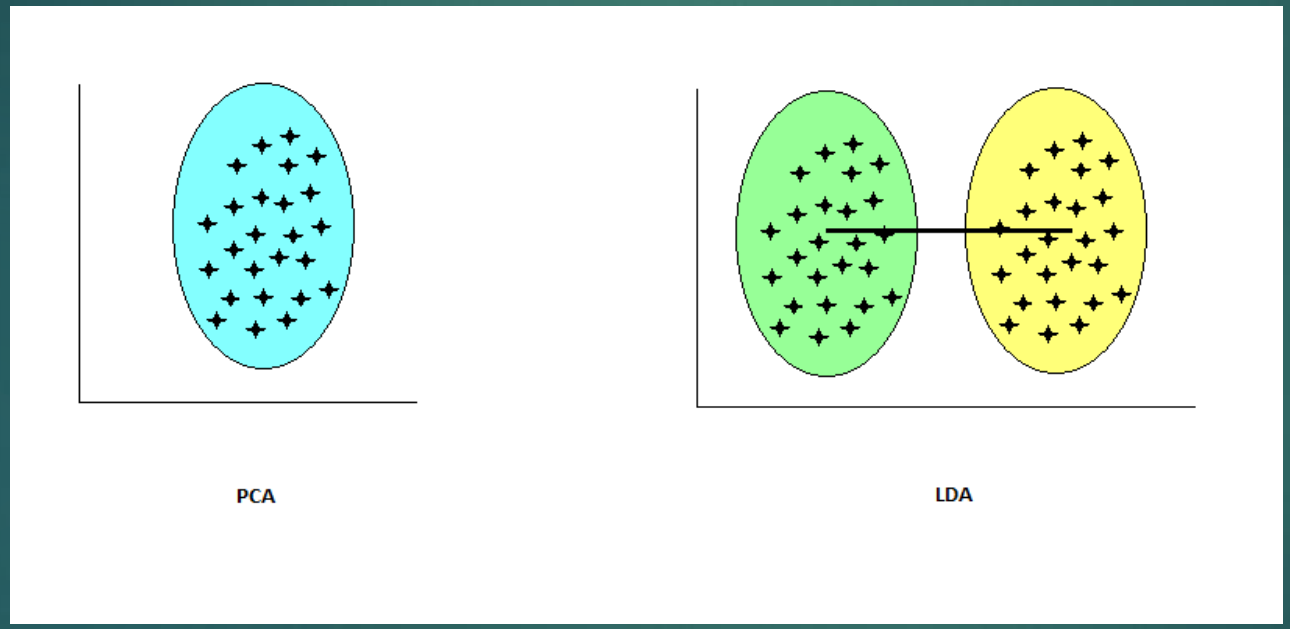
Classifiers Used

1. Linear Discriminant Analysis
2. Naive Bayes
3. K-Nearest neighbors
4. Support Vector Machines
5. Decision tree learning

Linear Discriminant Analysis (LDA)

- LDA is used to reduce the dimensions of a data set into lower dimensions.
- This new feature space will have better class label separation.
- LDA is a supervised algorithm because it takes in consideration of class labels of data.
- Whereas, PCA is unsupervised algorithm.

LDA vs PCA



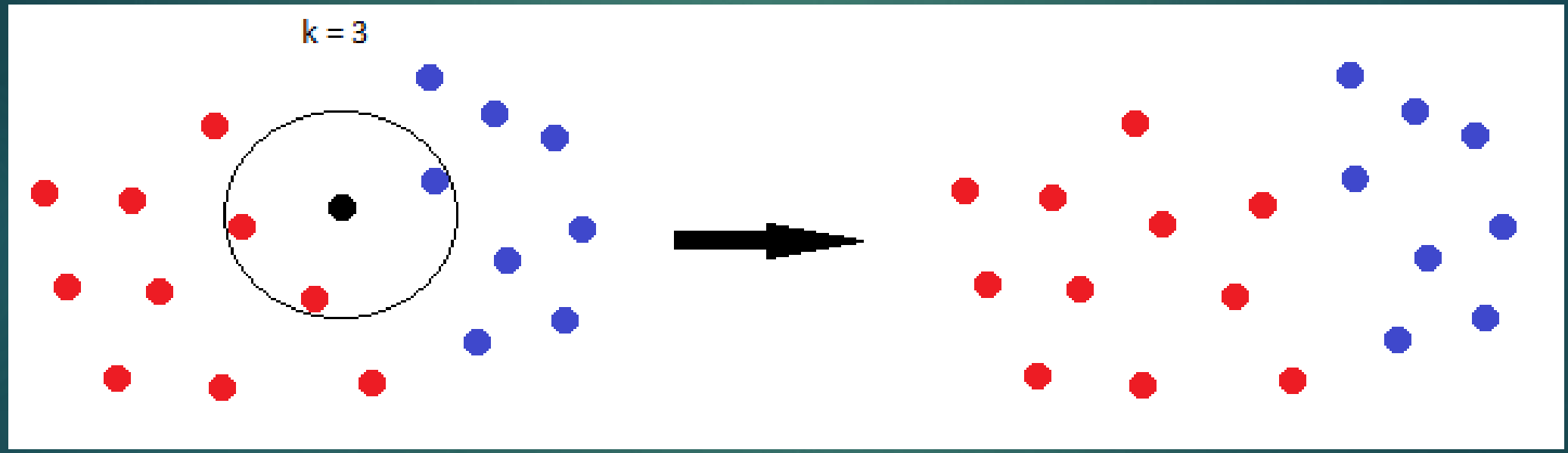
Naïve Bayes Classifier

- Simple
- Probabilistic classifier
- It uses Bayes Theorem.

K-Nearest neighbors

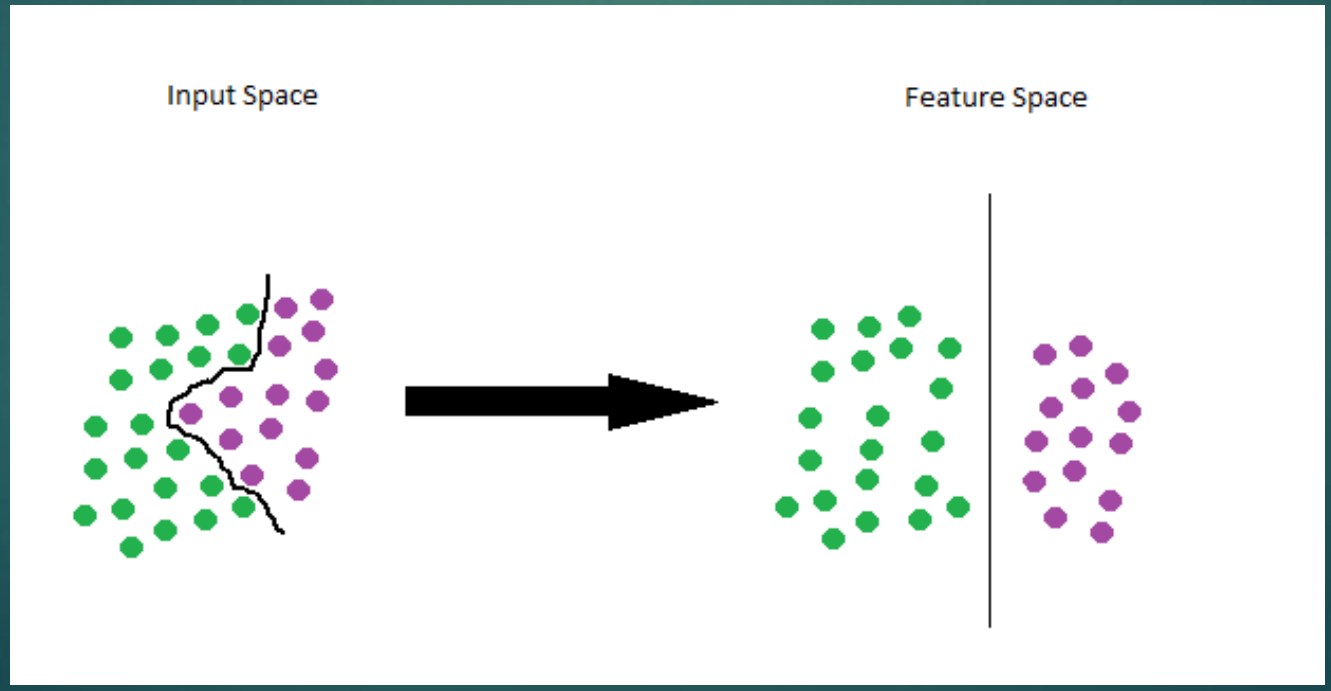
- K -Nearest neighbor algorithm has 2 steps.
  1. Training phase - Store the data and class labels.
  2. Algorithm is given two inputs,
     1. the value of k
     2. new data to be classified.
     - Then we find the k closest neighbors of the new data.
- The new data is then assigned a class label . How?
  - Whatever Class label had maximum occurrences in the k closest neighbors

K-Nearest neighbors



Support Vector Machines

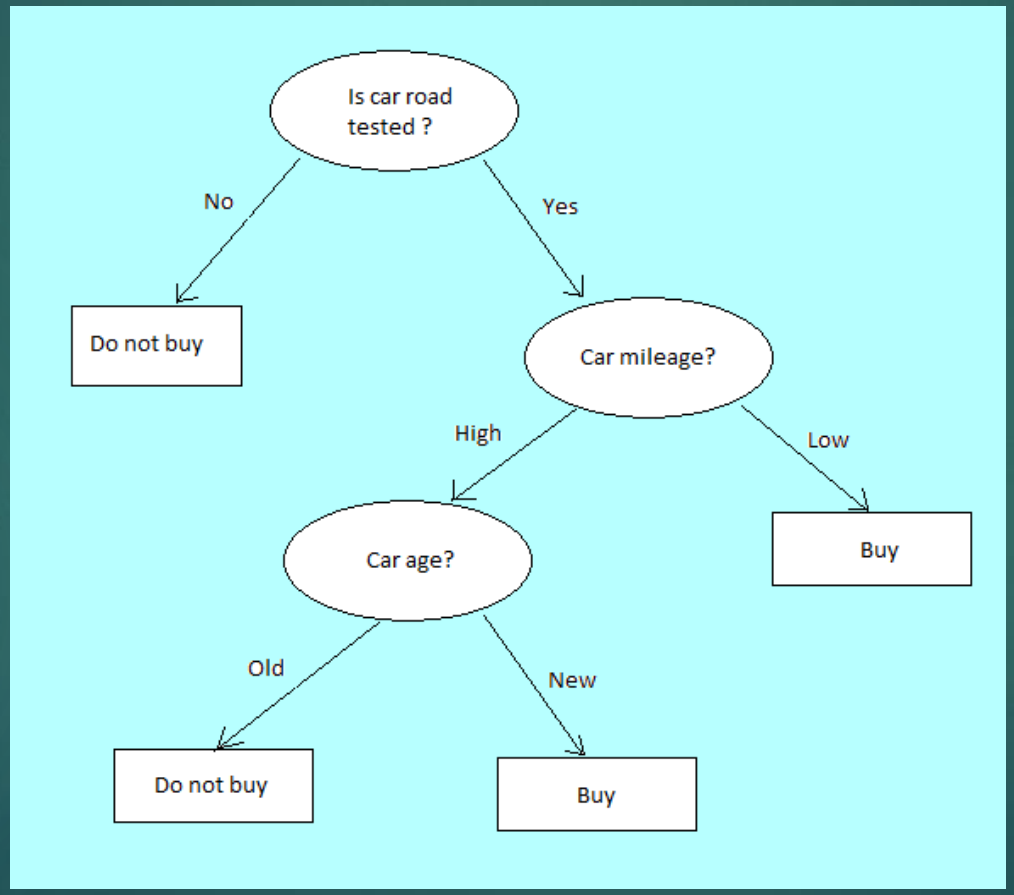
- In SVM, input space is mapped to feature space using kernels (functions).
- We can see that in feature space, we can separate data groups using a straight line.



Decision Tree Learning

- Decision tree learning is a learning algorithm.
- It looks a lot like flow chart.
  - The internal nodes in this tree contains tests.
  - The branches if this tree are the outcome.
  - The leaves of this tree have the label of class.

Decision Tree Learning



Running Application Code

- Refer to Readme file
- Project demonstration (later)

Existing Approaches

- Age and gender classification can be done using convolutional neural networks.
  - Accuracy of 86.8 percent.
- Gender classification based on gait.
  - Accuracy of 97.9%
- Gender classification using Local Directional Pattern (LDP). Accuracy of 95.05
- Gender classification based on speech.
  - Accuracy of 98% and 95 for noisy speech.
  - Disadvantage – Have to speak closer to mic

Result (Confusion Matrix)

LDA		Predicted Label	
Actual Label		Male	Female
	Male	40	0
	Female	1	39

KNN		Predicted Label	
Actual Label		Male	Female
	Male	40	0
	Female	2	38

SVM		Predicted Label	
Actual Label		Male	Female
	Male	40	0
	Female	3	37

Naïve Bayes		Predicted Label	
Actual Label		Male	Female
	Male	38	2
	Female	6	34

Result (Confusion Matrix) continued…

Decision Tree		Predicted Label	
Actual Label		Male	Female
	Male	31	9
	Female	6	34

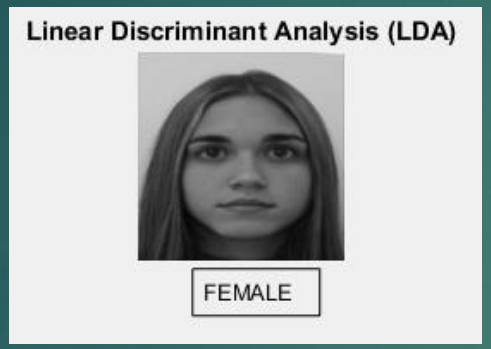
Results (Continued)…

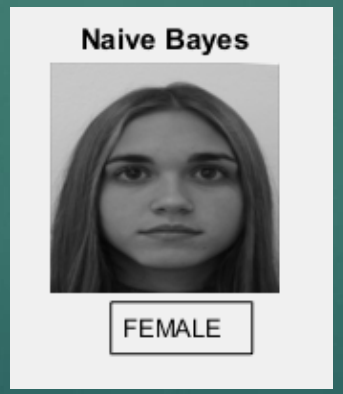
Classifier	Accuracy	Misclassification Rate (Error Rate)
Linear Discriminant Analysis	98.75%	1.25 %
K-Nearest neighbors	97.5%	2.5%
Support Vector Machines	96.25%	3.75%
Naive Bayes	90%	10%
Decision Tree Learning	81.25%	18.75%

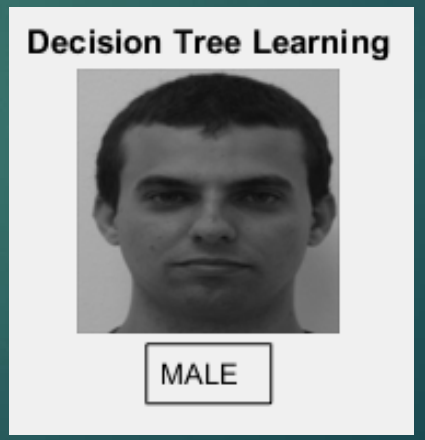
Results (Continued)…

Classifier	Precision	Specificity
Linear Discriminant Analysis	1	0.975
K-Nearest Neighbors	1	0.95
Support Vector Machines	1	0.925
Naive Bayes	0.95	0.85
Decision Tree Learning	0.775	0.85

Results (Screenshots)







Future Improvements

- Gender classification works very accurately for images.
- This project can be extended for video stream.
  - Since the time taken to predict is not that small, therefore output streaming would have some delay.

Conclusion

- Linear Discriminant Analysis, K-Nearest neighbors and Support Vector Machines have very good accuracy.
- In my project, LDA had highest accuracy among all the classifiers.
- LDA had only one wrong classification, KNN had 2 wrong classifications and SVM had 3 wrong classifications.
- Therefore, they can be used accurately for gender classification.
- Naive Bayes classification had 8 wrong classification with accuracy of 90%.
- Decision tree learning 15 wrong classifications with accuracy of 81.25%.
- Therefore these two algorithms are not suited for Gender Classification.

Project Demonstration !!!

Thanks for watching !!!