Machine Learning In Python
Subject : Classification Using QDA, Evaluating the model

Lecturer : Reza Akbari Movahed
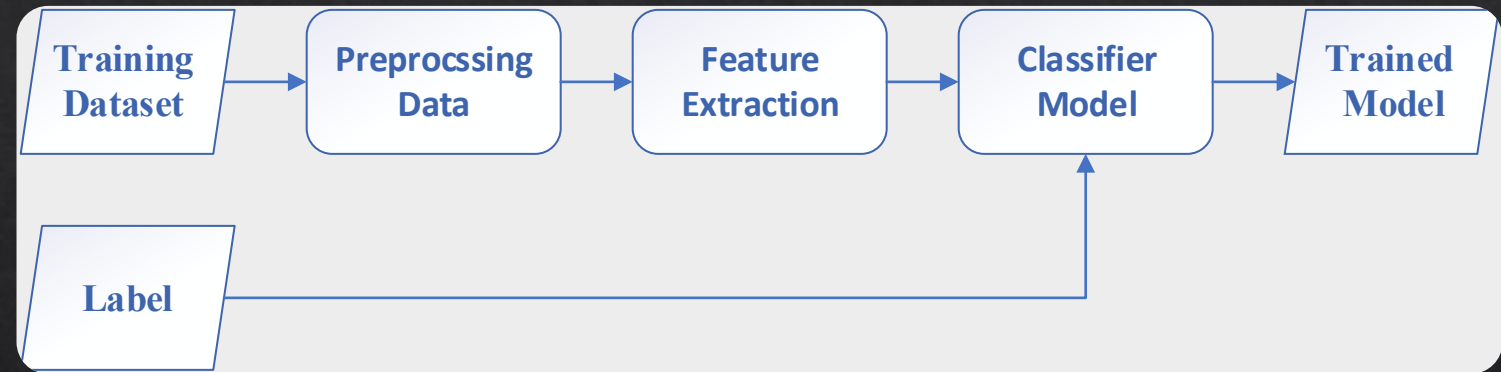
Hamedan University of Technology
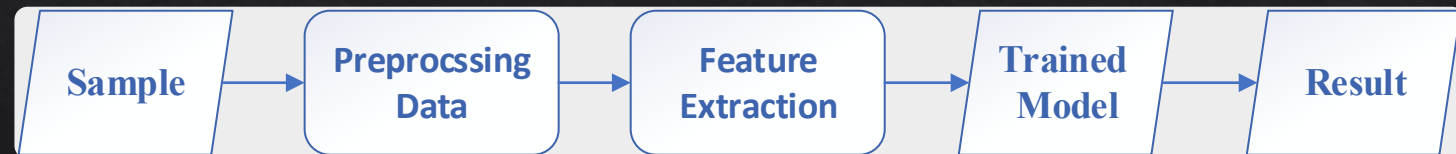
Winter 2020

Classification Using QDA

Classification In Supervised Learning Framework

Training Phase

Training Dataset → Preprocssing Data → Feature Extraction → Classifier Model → Trained Model

Label → Classifier Model

Testing Phase

Sample → Preprocssing Data → Feature Extraction → Trained Model → Result

Classification Using QDA
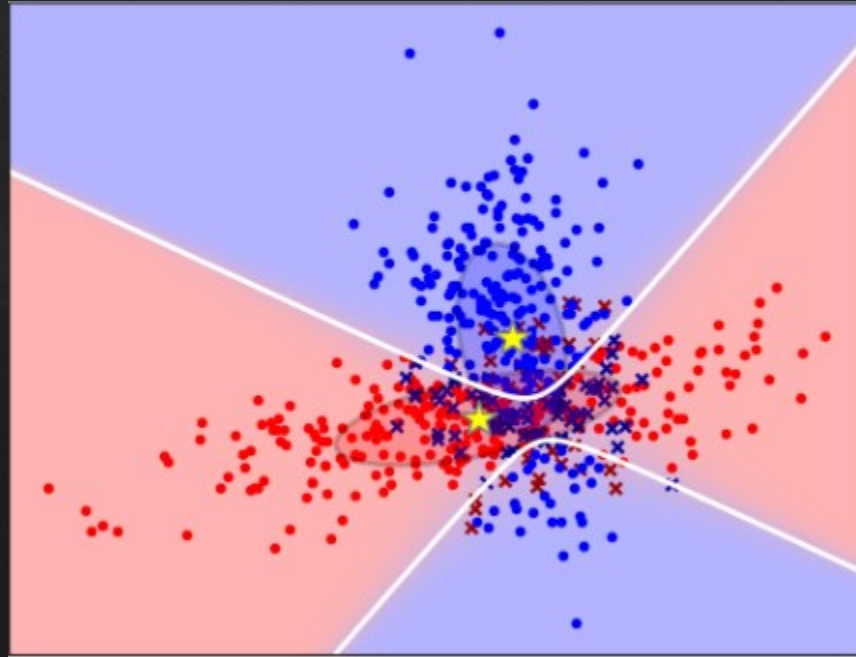
Classifier Models :

Quadratic Discriminant Analysis (QDA)

- An QDA classifier is a classifier that uses a quadratic decision surface to separate two or more classes.

- In the training stage, It finds an optimum quadratic surface with the most separability between classes

- It divides feature space into the different classes.

Classification Using QDA

Classifier Models :

Quadratic Discriminant Analysis (QDA)

Classification Using QDA

Breast cancer wisconsin (diagnostic) dataset

Number of Instances:	569
Number of Attributes:	30 numeric, predictive attributes and the class

Attribute Information:	• radius (mean of distances from center to points on the perimeter)
	• texture (standard deviation of gray-scale values)
	• perimeter
	• area
	• smoothness (local variation in radius lengths)
	• compactness (perimeter^2 / area - 1.0)
	• concavity (severity of concave portions of the contour)
	• concave points (number of concave portions of the contour)
	• symmetry
	• fractal dimension ("coastline approximation" - 1)
	
	The mean, standard error, and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.
	
	• class:
	    ○ WDBC-Malignant
	    ○ WDBC-Benign

Classification Using QDA

Breast cancer wisconsin (diagnostic) dataset

This is a copy of UCI ML Breast Cancer Wisconsin (Diagnostic) datasets. https://goo.gl/U2Uwz2

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image.

Separating plane described above was obtained using Multisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992], a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes.

The actual linear program used to obtain the separating plane in the 3-dimensional space is that described in: [K. P. Bennett and O. L. Mangasarian: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34].

This database is also available through the UW CS ftp server:

ftp ftp.cs.wisc.edu cd math-prog/cpo-dataset/machine-learn/WDBC/

Evaluating the model

$$Accuracy = \frac{N_t}{N}$$

$$Loss = 1 - Accuracy$$

N is the number

N_t is the number of samples that are correctly classified