

What is the concept of Quadratic Discriminant Analysis?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the concept of Quadratic Discriminant Analysis?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137803>

Quadratic Discriminant Analysis (QDA) is a statistical classification technique used to classify data into multiple groups based on a set of input variables. It is a type of discriminant analysis that assumes that the distributions of the input variables for each group follow a multivariate normal distribution, and that the covariance matrices for each group may be different. QDA uses quadratic functions to create decision boundaries between the groups, allowing for more flexibility in the shape of the decision boundary compared to Linear Discriminant Analysis. This method is commonly used for pattern recognition, image classification, and feature extraction, and is a powerful tool for predicting group membership based on a combination of input variables.

Introduction to Quadratic Discriminant Analysis

When we have a set of predictor variables and we'd like to classify a response variable into one of two classes, we typically use logistic regression.

However, when a response variable has more than two possible classes then we typically use linear discriminant analysis, often referred to as LDA.

LDA assumes that (1) observations from each class are normally distributed and (2) observations from each class share the same covariance matrix. Using these assumptions, LDA then finds the following values:

μ_k : The mean of all training observations from the k th class.
 σ^2 : The weighted average of the sample variances for each of the k classes.
 π_k : The proportion of the training observations that belong to the k th class.

LDA then plugs these numbers into the following formula and assigns each observation $X = x$ to the class for which the formula produces the largest value:

$$D_k(x) = x * (\mu_k/\sigma^2) - (\mu_k^2/2\sigma^2) + \log(\pi_k)$$

LDA has *linear* in its name because the value produced by the function above comes from a result of *linear functions* of x .

An extension of linear discriminant analysis is quadratic discriminant analysis, often referred to as QDA.

This method is similar to LDA and also assumes that the observations from each class are normally distributed, but it does not assume that each class shares the same covariance matrix. Instead, QDA assumes that each class has its own covariance matrix.

That is, it assumes that an observation from the k th class is of the form $X \sim N(\mu_k, \Sigma_k)$.

Using this assumption, QDA then finds the following values:

μ_k : The mean of all training observations from the k th class.
 Σ_k : The covariance matrix of the k th class.
 π_k : The proportion of the training observations that belong to the k th class.

QDA then plugs these numbers into the following formula and assigns each observation $X = x$ to the class for which the formula produces the largest value:

$$D_k(x) = -1/2*(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) - 1/2*\log|\Sigma_k| + \log(\pi_k)$$

Note that QDA has *quadratic* in its name because the value produced by the function above comes from a result of *quadratic functions* of x .

LDA vs. QDA: When to Use One vs. the Other

The main difference between LDA and QDA is that LDA assumes each class shares a covariance matrix, which makes it a much less flexible classifier than QDA.

QDA is generally preferred to LDA in the following situations:

(1) The training set is large.

(2) It's unlikely that the K classes share a common

covariance matrix.

When these conditions hold, QDA tends to perform better since it is more flexible and can provide a better fit to the data.

How to Prepare Data for QDA

Make sure your data meets the following requirements before applying a QDA model to it:

- 1. The response variable is categorical. QDA models are designed to be used for classification problems, i.e. when the response variable can be placed into classes or categories.**
- 2. The observations in each class follow a normal distribution. First, check that each the distribution of values in each class is roughly normally distributed. If this is not the case, you may choose to first transform the data to make the distribution more normal.**
- 3. Account for extreme outliers. Be sure to check for extreme outliers in the dataset before applying LDA. Typically you can check for outliers visually by simply using boxplots or scatterplots.**

QDA in R & Python

The following tutorials provide step-by-step examples of how to perform quadratic discriminant analysis in R and Python:

[Quadratic Discriminant Analysis in R \(Step-by-Step\)](#)

[Quadratic Discriminant Analysis in Python \(Step-by-Step\)](#)

ARABPSYCHOLOGY.COM