

# Linear discriminant analysis

---

**Linear discriminant analysis** (LDA), **normal discriminant analysis** (NDA), or **discriminant function analysis** is a generalization of **Fisher's linear discriminant**, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

LDA is closely related to analysis of variance (ANOVA) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements.<sup>[1][2]</sup> However, ANOVA uses categorical independent variables and a continuous dependent variable, whereas discriminant analysis has continuous independent variables and a categorical dependent variable (*i.e.* the class label).<sup>[3]</sup> Logistic regression and probit regression are more similar to LDA than ANOVA is, as they also explain a categorical variable by the values of continuous independent variables. These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the LDA method.

LDA is also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data.<sup>[4]</sup> LDA explicitly attempts to model the difference between the classes of data. PCA, in contrast, does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made.

LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.<sup>[5][6]</sup>

Discriminant analysis is used when groups are known a priori (unlike in cluster analysis). Each case must have a score on one or more quantitative predictor measures, and a score on a group measure.<sup>[7]</sup> In simple terms, discriminant function analysis is classification - the act of distributing things into groups, classes or categories of the same type.

## Contents

---

### History

### LDA for two classes

### Assumptions

### Discriminant functions

### Discrimination rules

### Eigenvalues

### Effect size

### Canonical discriminant analysis for *k* classes

## Fisher's linear discriminant

### Multiclass LDA

### Incremental LDA

### Practical use

### Applications

Bankruptcy prediction

Face recognition

Marketing

Biomedical studies

Earth science

### Comparison to logistic regression

### Linear discriminant in high dimension

### See also

### References

### Further reading

### External links

## History

---

The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936.<sup>[8]</sup> It is different from an ANOVA or MANOVA, which is used to predict one (ANOVA) or multiple (MANOVA) continuous dependent variables by one or more independent categorical variables. Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership.<sup>[9]</sup>

## LDA for two classes

---

Consider a set of observations  $\vec{x}$  (also called features, attributes, variables or measurements) for each sample of an object or event with known class  $y$ . This set of samples is called the training set. The classification problem is then to find a good predictor for the class  $y$  of any sample of the same distribution (not necessarily from the training set) given only an observation  $\vec{x}$ .<sup>[10]:338</sup>

LDA approaches the problem by assuming that the conditional probability density functions  $p(\vec{x}|y = 0)$  and  $p(\vec{x}|y = 1)$  are both the normal distribution with mean and covariance parameters  $(\vec{\mu}_0, \Sigma_0)$  and  $(\vec{\mu}_1, \Sigma_1)$ , respectively. Under this assumption, the Bayes optimal solution is to predict points as being from the second class if the log of the likelihood ratios is bigger than some threshold  $T$ , so that:

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x} - \vec{\mu}_0) + \ln |\Sigma_0| - (\vec{x} - \vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x} - \vec{\mu}_1) - \ln |\Sigma_1| > T$$

Without any further assumptions, the resulting classifier is referred to as quadratic discriminant analysis (QDA).

LDA instead makes the additional simplifying homoscedasticity assumption (*i.e.* that the class covariances are identical, so  $\Sigma_0 = \Sigma_1 = \Sigma$ ) and that the covariances have full rank. In this case, several terms cancel:

$$\vec{x}^T \Sigma_0^{-1} \vec{x} = \vec{x}^T \Sigma_1^{-1} \vec{x}$$

$$\vec{x}^T \Sigma_i^{-1} \vec{\mu}_i = \vec{\mu}_i^T \Sigma_i^{-1} \vec{x} \text{ because } \Sigma_i \text{ is Hermitian}$$

and the above decision criterion becomes a threshold on the dot product

$$\vec{w} \cdot \vec{x} > c$$

for some threshold constant  $c$ , where

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

$$c = \vec{w} \cdot \frac{1}{2}(\vec{\mu}_1 + \vec{\mu}_0)$$

This means that the criterion of an input  $\vec{x}$  being in a class  $y$  is purely a function of this linear combination of the known observations.

It is often useful to see this conclusion in geometrical terms: the criterion of an input  $\vec{x}$  being in a class  $y$  is purely a function of projection of multidimensional-space point  $\vec{x}$  onto vector  $\vec{w}$  (thus, we only consider its direction). In other words, the observation belongs to  $y$  if corresponding  $\vec{x}$  is located on a certain side of a hyperplane perpendicular to  $\vec{w}$ . The location of the plane is defined by the threshold  $c$ .

## Assumptions

---

The assumptions of discriminant analysis are the same as those for MANOVA. The analysis is quite sensitive to outliers and the size of the smallest group must be larger than the number of predictor variables.<sup>[7]</sup>

- Multivariate normality: Independent variables are normal for each level of the grouping variable.<sup>[9][7]</sup>
- Homogeneity of variance/covariance (homoscedasticity): Variances among group variables are the same across levels of predictors. Can be tested with Box's M statistic.<sup>[9]</sup> It has been suggested, however, that linear discriminant analysis be used when covariances are equal, and that quadratic discriminant analysis may be used when covariances are not equal.<sup>[7]</sup>
- Multicollinearity: Predictive power can decrease with an increased correlation between predictor variables.<sup>[7]</sup>
- Independence: Participants are assumed to be randomly sampled, and a participant's score on one variable is assumed to be independent of scores on that variable for all other participants.<sup>[9][7]</sup>

It has been suggested that discriminant analysis is relatively robust to slight violations of these assumptions,<sup>[11]</sup> and it has also been shown that discriminant analysis may still be reliable when using dichotomous variables (where multivariate normality is often violated).<sup>[12]</sup>

## Discriminant functions

---

Discriminant analysis works by creating one or more linear combinations of predictors, creating a new latent variable for each function. These functions are called discriminant functions. The number of functions possible is either  $N_g - 1$  where  $N_g$  = number of groups, or  $p$  (the number of predictors), whichever is smaller. The first function created maximizes the differences between groups on that function.

The second function maximizes differences on that function, but also must not be correlated with the previous function. This continues with subsequent functions with the requirement that the new function not be correlated with any of the previous functions.

Given group  $j$ , with  $\mathbb{R}_j$  sets of sample space, there is a discriminant rule such that if  $\mathbf{x} \in \mathbb{R}_j$ , then  $\mathbf{x} \in j$ . Discriminant analysis then, finds “good” regions of  $\mathbb{R}_j$  to minimize classification error, therefore leading to a high percent correct classified in the classification table.<sup>[13]</sup>

Each function is given a discriminant score to determine how well it predicts group placement.

- **Structure Correlation Coefficients:** The correlation between each predictor and the discriminant score of each function. This is a zero-order correlation (i.e., not corrected for the other predictors).<sup>[14]</sup>
- **Standardized Coefficients:** Each predictor's weight in the linear combination that is the discriminant function. Like in a regression equation, these coefficients are partial (i.e., corrected for the other predictors). Indicates the unique contribution of each predictor in predicting group assignment.
- **Functions at Group Centroids:** Mean discriminant scores for each grouping variable are given for each function. The farther apart the means are, the less error there will be in classification.

## Discrimination rules

---

- **Maximum likelihood:** Assigns  $\mathbf{x}$  to the group that maximizes population (group) density.<sup>[15]</sup>
- **Bayes Discriminant Rule:** Assigns  $\mathbf{x}$  to the group that maximizes  $\pi_i f_i(\mathbf{x})$ , where  $\pi_i$  represents the prior probability of that classification, and  $f_i(\mathbf{x})$  represents the population density.<sup>[15]</sup>
- **Fisher's linear discriminant rule:** Maximizes the ratio between  $SS_{\text{between}}$  and  $SS_{\text{within}}$ , and finds a linear combination of the predictors to predict group.<sup>[15]</sup>

## Eigenvalues

---

An eigenvalue in discriminant analysis is the characteristic root of each function. It is an indication of how well that function differentiates the groups, where the larger the eigenvalue, the better the function differentiates.<sup>[7]</sup> This however, should be interpreted with caution, as eigenvalues have no upper limit.<sup>[9][7]</sup> The eigenvalue can be viewed as a ratio of  $SS_{\text{between}}$  and  $SS_{\text{within}}$  as in ANOVA when the dependent variable is the discriminant function, and the groups are the levels of the IV.<sup>[9]</sup> This means that the largest eigenvalue is associated with the first function, the second largest with the second, etc..

## Effect size

---

Some suggest the use of eigenvalues as effect size measures, however, this is generally not supported.<sup>[9]</sup> Instead, the canonical correlation is the preferred measure of effect size. It is similar to the eigenvalue, but is the square root of the ratio of  $SS_{\text{between}}$  and  $SS_{\text{total}}$ . It is the correlation between groups and the function.<sup>[9]</sup> Another popular measure of effect size is the percent of variance for each function. This is calculated by:  $(\lambda_x / \Sigma \lambda_i) \times 100$  where  $\lambda_x$  is the eigenvalue for the function and  $\Sigma \lambda_i$  is the sum of all eigenvalues. This tells us how strong the prediction is for that particular function compared to the others.<sup>[9]</sup> Percent correctly

classified can also be analyzed as an effect size. The kappa value can describe this while correcting for chance agreement.<sup>[9]</sup> Kappa normalizes across all categories rather than biased by a significantly good or poorly performing classes.<sup>[16]</sup>

## Canonical discriminant analysis for $k$ classes

Canonical discriminant analysis (CDA) finds axes ( $k - 1$  canonical coordinates,  $k$  being the number of classes) that best separate the categories. These linear functions are uncorrelated and define, in effect, an optimal  $k - 1$  space through the  $n$ -dimensional cloud of data that best separates (the projections in that space of) the  $k$  groups. See “Multiclass LDA” for details below.

## Fisher's linear discriminant

The terms *Fisher's linear discriminant* and *LDA* are often used interchangeably, although Fisher's original article<sup>[1]</sup> actually describes a slightly different discriminant, which does not make some of the assumptions of LDA such as normally distributed classes or equal class covariances.

Suppose two classes of observations have means  $\vec{\mu}_0, \vec{\mu}_1$  and covariances  $\Sigma_0, \Sigma_1$ . Then the linear combination of features  $\vec{w} \cdot \vec{x}$  will have means  $\vec{w} \cdot \vec{\mu}_i$  and variances  $\vec{w}^T \Sigma_i \vec{w}$  for  $i = 0, 1$ . Fisher defined the separation between these two distributions to be the ratio of the variance between the classes to the variance within the classes:

$$S = \frac{\sigma_{\text{between}}^2}{\sigma_{\text{within}}^2} = \frac{(\vec{w} \cdot \vec{\mu}_1 - \vec{w} \cdot \vec{\mu}_0)^2}{\vec{w}^T \Sigma_1 \vec{w} + \vec{w}^T \Sigma_0 \vec{w}} = \frac{(\vec{w} \cdot (\vec{\mu}_1 - \vec{\mu}_0))^2}{\vec{w}^T (\Sigma_0 + \Sigma_1) \vec{w}}$$

This measure is, in some sense, a measure of the signal-to-noise ratio for the class labelling. It can be shown that the maximum separation occurs when

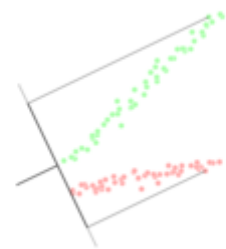
$$\vec{w} \propto (\Sigma_0 + \Sigma_1)^{-1} (\vec{\mu}_1 - \vec{\mu}_0)$$

When the assumptions of LDA are satisfied, the above equation is equivalent to LDA.

Be sure to note that the vector  $\vec{w}$  is the normal to the discriminant hyperplane. As an example, in a two dimensional problem, the line that best divides the two groups is perpendicular to  $\vec{w}$ .

Generally, the data points to be discriminated are projected onto  $\vec{w}$ ; then the threshold that best separates the data is chosen from analysis of the one-dimensional distribution. There is no general rule for the threshold. However, if projections of points from both classes exhibit approximately the same distributions, a good choice would be the hyperplane between projections of the two means,  $\vec{w} \cdot \vec{\mu}_0$  and  $\vec{w} \cdot \vec{\mu}_1$ . In this case the parameter  $c$  in threshold condition  $\vec{w} \cdot \vec{x} > c$  can be found explicitly:

$$c = \vec{w} \cdot \frac{1}{2} (\vec{\mu}_0 + \vec{\mu}_1) = \frac{1}{2} \vec{\mu}_1^T \Sigma_1^{-1} \vec{\mu}_1 - \frac{1}{2} \vec{\mu}_0^T \Sigma_0^{-1} \vec{\mu}_0.$$



Fisher's Linear Discriminant visualised as an axis

Otsu's method is related to Fisher's linear discriminant, and was created to binarize the histogram of pixels in a grayscale image by optimally picking the black/white threshold that minimizes intra-class variance and maximizes inter-class variance within/between grayscales assigned to black and white pixel classes.

## Multiclass LDA

In the case where there are more than two classes, the analysis used in the derivation of the Fisher discriminant can be extended to find a subspace which appears to contain all of the class variability.<sup>[17]</sup> This generalization is due to C. R. Rao.<sup>[18]</sup> Suppose that each of  $C$  classes has a mean  $\mu_i$  and the same covariance  $\Sigma$ . Then the scatter between class variability may be defined by the sample covariance of the class means

$$\Sigma_b = \frac{1}{C} \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T$$

where  $\mu$  is the mean of the class means. The class separation in a direction  $\vec{w}$  in this case will be given by

$$S = \frac{\vec{w}^T \Sigma_b \vec{w}}{\vec{w}^T \Sigma \vec{w}}$$

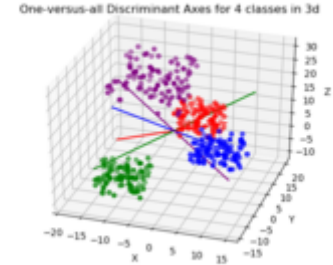
This means that when  $\vec{w}$  is an eigenvector of  $\Sigma^{-1} \Sigma_b$  the separation will be equal to the corresponding eigenvalue.

If  $\Sigma^{-1} \Sigma_b$  is diagonalizable, the variability between features will be contained in the subspace spanned by the eigenvectors corresponding to the  $C - 1$  largest eigenvalues (since  $\Sigma_b$  is of rank  $C - 1$  at most). These eigenvectors are primarily used in feature reduction, as in PCA. The eigenvectors corresponding to the smaller eigenvalues will tend to be very sensitive to the exact choice of training data, and it is often necessary to use regularisation as described in the next section.

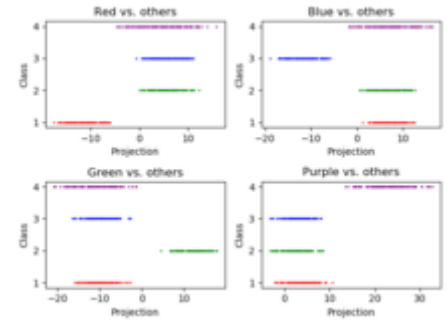
If classification is required, instead of dimension reduction, there are a number of alternative techniques available. For instance, the classes may be partitioned, and a standard Fisher discriminant or LDA used to classify each partition. A common example of this is "one against the rest" where the points from one class are put in one group, and everything else in the other, and then LDA applied. This will result in  $C$  classifiers, whose results are combined. Another common method is pairwise classification, where a new classifier is created for each pair of classes (giving  $C(C - 1)/2$  classifiers in total), with the individual classifiers combined to produce a final classification.

## Incremental LDA

The typical implementation of the LDA technique requires that all the samples are available in advance. However, there are situations where the entire data set is not available and the input data are observed as a stream. In this case, it is desirable for the LDA feature extraction to have the ability to update the computed LDA features by observing the new samples without running the algorithm on the whole data set. For example, in many real-time applications such as mobile robotics or on-line face recognition, it is important



Visualisation for one-versus-all LDA axes for 4 classes in 3d



Projections along linear discriminant axes for 4 classes

to update the extracted LDA features as soon as new observations are available. An LDA feature extraction technique that can update the LDA features by simply observing new samples is an *incremental LDA algorithm*, and this idea has been extensively studied over the last two decades.<sup>[19]</sup> Chatterjee and Roychowdhury proposed an incremental self-organized LDA algorithm for updating the LDA features.<sup>[20]</sup> In other work, Demir and Ozmehmet proposed online local learning algorithms for updating LDA features incrementally using error-correcting and the Hebbian learning rules.<sup>[21]</sup> Later, Aliyari *et al.* derived fast incremental algorithms to update the LDA features by observing the new samples.<sup>[19]</sup>

## Practical use

---

In practice, the class means and covariances are not known. They can, however, be estimated from the training set. Either the maximum likelihood estimate or the maximum a posteriori estimate may be used in place of the exact value in the above equations. Although the estimates of the covariance may be considered optimal in some sense, this does not mean that the resulting discriminant obtained by substituting these values is optimal in any sense, even if the assumption of normally distributed classes is correct.

Another complication in applying LDA and Fisher's discriminant to real data occurs when the number of measurements of each sample (i.e., the dimensionality of each data vector) exceeds the number of samples in each class.<sup>[4]</sup> In this case, the covariance estimates do not have full rank, and so cannot be inverted. There are a number of ways to deal with this. One is to use a pseudo inverse instead of the usual matrix inverse in the above formulae. However, better numeric stability may be achieved by first projecting the problem onto the subspace spanned by  $\Sigma_b$ .<sup>[22]</sup> Another strategy to deal with small sample size is to use a shrinkage estimator of the covariance matrix, which can be expressed mathematically as

$$\Sigma = (1 - \lambda)\Sigma + \lambda I$$

where  $I$  is the identity matrix, and  $\lambda$  is the *shrinkage intensity* or *regularisation parameter*. This leads to the framework of regularized discriminant analysis<sup>[23]</sup> or shrinkage discriminant analysis.<sup>[24]</sup>

Also, in many practical cases linear discriminants are not suitable. LDA and Fisher's discriminant can be extended for use in non-linear classification via the kernel trick. Here, the original observations are effectively mapped into a higher dimensional non-linear space. Linear classification in this non-linear space is then equivalent to non-linear classification in the original space. The most commonly used example of this is the kernel Fisher discriminant.

LDA can be generalized to multiple discriminant analysis, where  $c$  becomes a categorical variable with  $N$  possible states, instead of only two. Analogously, if the class-conditional densities  $p(\vec{x} | c = i)$  are normal with shared covariances, the sufficient statistic for  $P(c | \vec{x})$  are the values of  $N$  projections, which are the subspace spanned by the  $N$  means, affine projected by the inverse covariance matrix. These projections can be found by solving a generalized eigenvalue problem, where the numerator is the covariance matrix formed by treating the means as the samples, and the denominator is the shared covariance matrix. See “Multiclass LDA” above for details.

## Applications

---

In addition to the examples given below, LDA is applied in positioning and product management.

### Bankruptcy prediction

In bankruptcy prediction based on accounting ratios and other financial variables, linear discriminant analysis was the first statistical method applied to systematically explain which firms entered bankruptcy vs. survived. Despite limitations including known nonconformance of accounting ratios to the normal distribution assumptions of LDA, Edward Altman's 1968 model is still a leading model in practical applications.

## Face recognition

In computerised face recognition, each face is represented by a large number of pixel values. Linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template. The linear combinations obtained using Fisher's linear discriminant are called *Fisher faces*, while those obtained using the related principal component analysis are called *eigenfaces*.

## Marketing

In marketing, discriminant analysis was once often used to determine the factors which distinguish different types of customers and/or products on the basis of surveys or other forms of collected data. Logistic regression or other methods are now more commonly used. The use of discriminant analysis in marketing can be described by the following steps:

1. Formulate the problem and gather data—Identify the salient attributes consumers use to evaluate products in this category—Use quantitative marketing research techniques (such as surveys) to collect data from a sample of potential customers concerning their ratings of all the product attributes. The data collection stage is usually done by marketing research professionals. Survey questions ask the respondent to rate a product from one to five (or 1 to 7, or 1 to 10) on a range of attributes chosen by the researcher. Anywhere from five to twenty attributes are chosen. They could include things like: ease of use, weight, accuracy, durability, colourfulness, price, or size. The attributes chosen will vary depending on the product being studied. The same question is asked about all the products in the study. The data for multiple products is codified and input into a statistical program such as R, SPSS or SAS. (This step is the same as in Factor analysis).
2. Estimate the Discriminant Function Coefficients and determine the statistical significance and validity—Choose the appropriate discriminant analysis method. The direct method involves estimating the discriminant function so that all the predictors are assessed simultaneously. The stepwise method enters the predictors sequentially. The two-group method should be used when the dependent variable has two categories or states. The multiple discriminant method is used when the dependent variable has three or more categorical states. Use Wilks's Lambda to test for significance in SPSS or F stat in SAS. The most common method used to test validity is to split the sample into an estimation or analysis sample, and a validation or holdout sample. The estimation sample is used in constructing the discriminant function. The validation sample is used to construct a classification matrix which contains the number of correctly classified and incorrectly classified cases. The percentage of correctly classified cases is called the *hit ratio*.
3. Plot the results on a two dimensional map, define the dimensions, and interpret the results. The statistical program (or a related module) will map the results. The map will plot each product (usually in two-dimensional space). The distance of products to each other indicate either how different they are. The dimensions must be labelled by the researcher. This requires subjective judgement and is often very challenging. See perceptual mapping.



## Biomedical studies

The main application of discriminant analysis in medicine is the assessment of severity state of a patient and prognosis of disease outcome. For example, during retrospective analysis, patients are divided into groups according to severity of disease – mild, moderate and severe form. Then results of clinical and laboratory analyses are studied in order to reveal variables which are statistically different in studied groups. Using these variables, discriminant functions are built which help to objectively classify disease in a future patient into mild, moderate or severe form.

In biology, similar principles are used in order to classify and define groups of different biological objects, for example, to define phage types of *Salmonella enteritidis* based on Fourier transform infrared spectra,<sup>[25]</sup> to detect animal source of *Escherichia coli* studying its virulence factors<sup>[26]</sup> etc.

## Earth science

This method can be used to separate the alteration zones. For example, when different data from various zones are available, discriminant analysis can find the pattern within the data and classify it effectively.<sup>[27]</sup>

## Comparison to logistic regression

---

Discriminant function analysis is very similar to logistic regression, and both can be used to answer the same research questions.<sup>[9]</sup> Logistic regression does not have as many assumptions and restrictions as discriminant analysis. However, when discriminant analysis' assumptions are met, it is more powerful than logistic regression.<sup>[28]</sup> Unlike logistic regression, discriminant analysis can be used with small sample sizes. It has been shown that when sample sizes are equal, and homogeneity of variance/covariance holds, discriminant analysis is more accurate.<sup>[7]</sup> Despite all these advantages, logistic regression has none-the-less become the common choice, since the assumptions of discriminant analysis are rarely met.<sup>[8][7]</sup>

## Linear discriminant in high dimension

---

Geometric anomalies in higher dimensions lead to the well-known curse of dimensionality. Nevertheless, proper utilization of concentration of measure phenomena can make computation easier.<sup>[29]</sup> An important case of these *blissing of dimensionality* phenomena was highlighted by Donoho and Tanner: if a sample is essentially high-dimensional then each point can be separated from the rest of the sample by linear inequality, with high probability, even for exponentially large samples.<sup>[30]</sup> These linear inequalities can be selected in the standard (Fisher's) form of the linear discriminant for a rich family of probability distribution.<sup>[31]</sup> In particular, such theorems are proven for log-concave distributions including multidimensional normal distribution (the proof is based on the concentration inequalities for log-concave measures<sup>[32]</sup>) and for product measures on a multidimensional cube (this is proven using Talagrand's concentration inequality for product probability spaces). Data separability by classical linear discriminants simplifies the problem of error correction for artificial intelligence systems in high dimension.<sup>[33]</sup>

## See also

---

- Data mining
- Decision tree learning
- Factor analysis
- Kernel Fisher discriminant analysis

- Logit (for logistic regression)
- Linear regression
- Multiple discriminant analysis
- Multidimensional scaling
- Pattern recognition
- Preference regression
- Quadratic classifier
- Statistical classification

## References

---

1. Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems" (<https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15227/1/138.pdf>) (PDF). *Annals of Eugenics*. 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x (<https://doi.org/10.1111/1%2Fj.1469-1809.1936.tb02137.x>). hdl:2440/15227 (<https://hdl.handle.net/2440%2F15227>).
2. McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience. ISBN 978-0-471-69115-0. MR 1190469 (<https://www.ams.org/mathscinet-getitem?mr=1190469>).
3. Analyzing Quantitative Data: An Introduction for Social Researchers, Debra Wetcher-Hendricks, p.288
4. Martinez, A. M.; Kak, A. C. (2001). "PCA versus LDA" (<http://www.ece.osu.edu/~aleix/pami01.pdf>) (PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 23 (2): 228–233. doi:10.1109/34.908974 (<https://doi.org/10.1109%2F34.908974>).
5. Abdi, H. (2007) "Discriminant correspondence analysis." (<http://www.utdallas.edu/~herve/Abdi-DCA2007-pretty.pdf>) In: N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistic*. Thousand Oaks (CA): Sage. pp. 270–275.
6. Perriere, G.; Thioulouse, J. (2003). "Use of Correspondence Discriminant Analysis to predict the subcellular location of bacterial proteins". *Computer Methods and Programs in Biomedicine*. 70 (2): 99–105. doi:10.1016/s0169-2607(02)00011-1 (<https://doi.org/10.1016%2Fs0169-2607%2802%2900011-1>). PMID 12507786 (<https://pubmed.ncbi.nlm.nih.gov/12507786>).
7. BÖKEOĞLU ÇOKLUK, Ö, & BÜYÜKÖZTÜRK, Ş. (2008). Discriminant function analysis: Concept and application (<https://web.archive.org/web/20190607172058/https://pdfs.semanticscholar.org/76b1/c725b926d91db6eb52e83ab1596e6eb115ca.pdf>). Eğitim araştırmaları dergisi, (33), 73-92.
8. Cohen et al. Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences 3rd ed. (2003). Taylor & Francis Group.
9. Green, S.B. Salkind, N. J. & Akey, T. M. (2008). Using SPSS for Windows and Macintosh: Analyzing and understanding data (<https://www.tandfonline.com/doi/pdf/10.1198/tas.2005.s139>). New Jersey: Prentice Hall.
10. Venables, W. N.; Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer Verlag. ISBN 978-0-387-95457-8.
11. Lachenbruch, P. A. (1975). *Discriminant analysis*. NY: Hafner
12. Klecka, William R. (1980). *Discriminant analysis*. Quantitative Applications in the Social Sciences Series, No. 19. Thousand Oaks, CA: Sage Publications.
13. Hardle, W., Simar, L. (2007). *Applied Multivariate Statistical Analysis*. Springer Berlin Heidelberg. pp. 289–303.
14. Garson, G. D. (2008). Discriminant function analysis. <https://web.archive.org/web/20080312065328/http://www2.chass.ncsu.edu/garson/pA765/disc>

15. Hardle, W., Simar, L. (2007). *Applied Multivariate Statistical Analysis* (<https://pdfs.semanticscholar.org/e49a/3ed1560a0979f5ca0146c1aca7ef98e64af7.pdf>). Springer Berlin Heidelberg. pp. 289-303.
16. Israel, Steven A. (June 2006). "Performance Metrics: How and When". *Geocarto International*. **21** (2): 23–32. doi:10.1080/10106040608542380 (<https://doi.org/10.1080%2F10106040608542380>). ISSN 1010-6049 (<https://www.worldcat.org/issn/1010-6049>). S2CID 122376081 (<https://api.semanticscholar.org/CorpusID:122376081>).
17. Garson, G. D. (2008). Discriminant function analysis. "Archived copy" (<https://web.archive.org/web/20080312065328/http://www2.chass.ncsu.edu/garson/pA765/discrim.htm>). Archived from the original (<http://www2.chass.ncsu.edu/garson/pa765/discrim.htm>) on 2008-03-12. Retrieved 2008-03-04. .
18. Rao, R. C. (1948). "The utilization of multiple measurements in problems of biological classification". *Journal of the Royal Statistical Society, Series B*. **10** (2): 159–203. JSTOR 2983775 (<https://www.jstor.org/stable/2983775>).
19. Aliyari Ghassabeh, Youness; Rudzicz, Frank; Moghaddam, Hamid Abrishami (2015-06-01). "Fast incremental LDA feature extraction". *Pattern Recognition*. **48** (6): 1999–2012. Bibcode:2015PatRe..48.1999A (<https://ui.adsabs.harvard.edu/abs/2015PatRe..48.1999A>). doi:10.1016/j.patcog.2014.12.012 (<https://doi.org/10.1016%2Fj.patcog.2014.12.012>).
20. Chatterjee, C.; Roychowdhury, V.P. (1997-05-01). "On self-organizing algorithms and networks for class-separability features". *IEEE Transactions on Neural Networks*. **8** (3): 663–678. doi:10.1109/72.572105 (<https://doi.org/10.1109%2F72.572105>). ISSN 1045-9227 (<http://www.worldcat.org/issn/1045-9227>). PMID 18255669 (<https://pubmed.ncbi.nlm.nih.gov/18255669>).
21. Demir, G. K.; Ozmehmet, K. (2005-03-01). "Online Local Learning Algorithms for Linear Discriminant Analysis". *Pattern Recognit. Lett.* **26** (4): 421–431. Bibcode:2005PaReL..26..421D (<https://ui.adsabs.harvard.edu/abs/2005PaReL..26..421D>). doi:10.1016/j.patrec.2004.08.005 (<https://doi.org/10.1016%2Fj.patrec.2004.08.005>). ISSN 0167-8655 (<https://www.worldcat.org/issn/0167-8655>).
22. Yu, H.; Yang, J. (2001). "A direct LDA algorithm for high-dimensional data — with application to face recognition". *Pattern Recognition*. **34** (10): 2067–2069. Bibcode:2001PatRe..34.2067Y (<https://ui.adsabs.harvard.edu/abs/2001PatRe..34.2067Y>). CiteSeerX 10.1.1.70.3507 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.70.3507>). doi:10.1016/s0031-3203(00)00162-x (<https://doi.org/10.1016%2Fs0031-3203%2800%2900162-x>).
23. Friedman, J. H. (1989). "Regularized Discriminant Analysis" (<http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-4389.pdf>) (PDF). *Journal of the American Statistical Association*. **84** (405): 165–175. CiteSeerX 10.1.1.382.2682 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.382.2682>). doi:10.2307/2289860 (<https://doi.org/10.2307%2F2289860>). JSTOR 2289860 (<https://www.jstor.org/stable/2289860>). MR 0999675 (<https://www.ams.org/mathscinet-getitem?mr=0999675>).
24. Ahdesmäki, M.; Strimmer, K. (2010). "Feature selection in omics prediction problems using cat scores and false nondiscovery rate control". *Annals of Applied Statistics*. **4** (1): 503–519. arXiv:0903.2003 (<https://arxiv.org/abs/0903.2003>). doi:10.1214/09-aos277 (<https://doi.org/10.1214%2F09-aos277>). S2CID 2508935 (<https://api.semanticscholar.org/CorpusID:2508935>).

25. Preisner, O; Guiomar, R; Machado, J; Menezes, JC; Lopes, JA (2010). "Application of Fourier transform infrared spectroscopy and chemometrics for differentiation of *Salmonella enterica* serovar Enteritidis phage types" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2876429>). *Appl Environ Microbiol.* **76** (11): 3538–3544. Bibcode:2010ApEnM..76.3538P (<http://ui.adsabs.harvard.edu/abs/2010ApEnM..76.3538P>). doi:10.1128/aem.01589-09 (<https://doi.org/10.1128%2Faem.01589-09>). PMC 2876429 (<https://www.ncbi.nlm.nih.gov/pmc/article/s/PMC2876429>). PMID 20363777 (<https://pubmed.ncbi.nlm.nih.gov/20363777>).
26. David, DE; Lynne, AM; Han, J; Foley, SL (2010). "Evaluation of virulence factor profiling in the characterization of veterinary *Escherichia coli* isolates" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976202>). *Appl Environ Microbiol.* **76** (22): 7509–7513. Bibcode:2010ApEnM..76.7509D (<https://ui.adsabs.harvard.edu/abs/2010ApEnM..76.7509D>). doi:10.1128/aem.00726-10 (<https://doi.org/10.1128%2Faem.00726-10>). PMC 2976202 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976202>). PMID 20889790 (<https://pubmed.ncbi.nlm.nih.gov/20889790>).
27. Tahmasebi, P.; Hezarkhani, A.; Mortazavi, M. (2010). "Application of discriminant analysis for alteration separation; sungun copper deposit, East Azerbaijan, Iran. Australian" (<http://ajbasweb.com/old/ajbas/2010/564-576.pdf>) (PDF). *Journal of Basic and Applied Sciences.* **6** (4): 564–576.
28. Trevor Hastie; Robert Tibshirani; Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction* (second ed.). Springer. p. 128.
29. Kainen P.C. (1997) Utilizing geometric anomalies of high dimension: When complexity makes computation easier (<https://web.archive.org/web/20190226172352/http://pdfs.semanticscholar.org/708f/8e0a95ba5977072651c0681f3c7b8f09eca3.pdf>). In: Kárný M., Warwick K. (eds) *Computer Intensive Methods in Control and Signal Processing: The Curse of Dimensionality*, Springer, 1997, pp. 282–294.
30. Donoho, D., Tanner, J. (2009) Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing (<https://arxiv.org/abs/0906.2530>), *Phil. Trans. R. Soc. A* **367**, 4273–4293.
31. Gorban, Alexander N.; Golubkov, Alexander; Grechuck, Bogdan; Mirkes, Evgeny M.; Tyukin, Ivan Y. (2018). "Correction of AI systems by linear discriminants: Probabilistic foundations". *Information Sciences.* **466**: 303–322. arXiv:1811.05321 (<https://arxiv.org/abs/1811.05321>). doi:10.1016/j.ins.2018.07.040 (<https://doi.org/10.1016%2Fj.ins.2018.07.040>). S2CID 52876539 (<https://api.semanticscholar.org/CorpusID:52876539>).
32. Guédon, O., Milman, E. (2011) Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures (<https://arxiv.org/abs/1011.0943>), *Geom. Funct. Anal.* **21** (5), 1043–1068.
33. Gorban, Alexander N.; Makarov, Valeri A.; Tyukin, Ivan Y. (July 2019). "The unreasonable effectiveness of small neural ensembles in high-dimensional brain" (<https://doi.org/10.1016%2Fj.pprev.2018.09.005>). *Physics of Life Reviews.* **29**: 55–88. arXiv:1809.07656 (<https://arxiv.org/abs/1809.07656>). Bibcode:2019PhLRv..29...55G (<https://ui.adsabs.harvard.edu/abs/2019PhLRv..29...55G>). doi:10.1016/j.pprev.2018.09.005 (<https://doi.org/10.1016%2Fj.pprev.2018.09.005>). PMID 30366739 (<https://pubmed.ncbi.nlm.nih.gov/30366739>).

## Further reading

---

- Duda, R. O.; Hart, P. E.; Stork, D. H. (2000). *Pattern Classification* (2nd ed.). Wiley Interscience. ISBN 978-0-471-05669-0. MR 1802993 (<https://www.ams.org/mathscinet-getitem?mr=1802993>).
- Hilbe, J. M. (2009). *Logistic Regression Models*. Chapman & Hall/CRC Press. ISBN 978-1-4200-7575-5.

- Mika, S.; et al. (1999). "Fisher Discriminant Analysis with Kernels". *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*. *IEEE Conference on Neural Networks for Signal Processing IX*. pp. 41–48. CiteSeerX 10.1.1.35.9904 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.9904>). doi:10.1109/NNSP.1999.788121 (<https://doi.org/10.1109%2FNNSP.1999.788121>). ISBN 978-0-7803-5673-3. S2CID 8473401 (<https://api.semanticscholar.org/CorpusID:8473401>).
- McFarland, H. Richard; Donald, St. P. Richards (2001). "Exact Misclassification Probabilities for Plug-In Normal Quadratic Discriminant Functions. I. The Equal-Means Case" (<https://doi.org/10.1006%2Fjmva.2000.1924>). *Journal of Multivariate Analysis*. **77** (1): 21–53. doi:10.1006/jmva.2000.1924 (<https://doi.org/10.1006%2Fjmva.2000.1924>).
- McFarland, H. Richard; Donald, St. P. Richards (2002). "Exact Misclassification Probabilities for Plug-In Normal Quadratic Discriminant Functions. II. The Heterogeneous Case" (<https://doi.org/10.1006%2Fjmva.2001.2034>). *Journal of Multivariate Analysis*. **82** (2): 299–330. doi:10.1006/jmva.2001.2034 (<https://doi.org/10.1006%2Fjmva.2001.2034>).
- Haghighat, M.; Abdel-Mottaleb, M.; Alhalabi, W. (2016). "Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition" (<https://zenodo.org/record/889881>). *IEEE Transactions on Information Forensics and Security*. **11** (9): 1984–1996. doi:10.1109/TIFS.2016.2569061 (<https://doi.org/10.1109%2FTIFS.2016.2569061>). S2CID 15624506 (<https://api.semanticscholar.org/CorpusID:15624506>).

## External links

---

- Discriminant Correlation Analysis (DCA) of the Haghighat article (see above) (<https://github.com/mhaghighat/dcaFuse>)
  - ALGLIB (<http://www.alglib.net/dataanalysis/lineardiscriminantanalysis.php>) contains open-source LDA implementation in C# / C++ / Pascal / VBA.
  - LDA in Python (<https://www.mltut.com/linear-discriminant-analysis-python-complete-and-easy-guide/>)- LDA implementation in Python
  - LDA tutorial using MS Excel (<http://people.revoledu.com/kardi/tutorial/LDA/index.html>)
  - Biomedical statistics. Discriminant analysis (<https://web.archive.org/web/20150405124836/http://biostat.katernyakon.in.ua/en/prognosis/discriminant-analysis.html>)
  - StatQuest: Linear Discriminant Analysis (LDA) clearly explained (<https://www.youtube.com/watch?v=azXCzI57Yfc>) on YouTube
  - Course notes, Discriminant function analysis by G. David Garson, NC State University (<http://web.archive.org/web/20080312065328/http://www2.chass.ncsu.edu/garson/pA765/discrim.htm>)
  - Discriminant analysis tutorial in Microsoft Excel by Kardi Teknomo (<http://people.revoledu.com/kardi/tutorial/LDA/>)
  - Course notes, Discriminant function analysis by David W. Stockburger, Missouri State University (<http://www.psychstat.missouristate.edu/multibook/mlt03m.html>)
  - Discriminant function analysis (DA) by John Poulsen and Aaron French, San Francisco State University (<http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf>)
- 

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Linear\\_discriminant\\_analysis&oldid=1095695170](https://en.wikipedia.org/w/index.php?title=Linear_discriminant_analysis&oldid=1095695170)"

---

This page was last edited on 29 June 2022, at 20:19 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License 3.0; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

