Package: lyubishchev (via r-universe)

June 22, 2026

Type Package

Title Quantitative Taxonomy Methods of A.A. Lyubishchev (1943)

Version 0.1.0

Description Implements the multivariate classification methods of Alexander Alexandrovich Lyubishchev (1890-1972), as described in his 1943 manuscript 'Programma obshchey sistematiki' Lyubishchev (1943) <https://www.zin.ru/animalia/coleoptera/rus/lyubis05.htm> and published in Lubischew (1962) <https://www.jstor.org/stable/2527894>. Provides divergence_coefficient() for measuring separation between groups on continuous features, scatter_ellipse() for fitting covariance ellipses per class, transgression() for detecting ellipse overlap, and classify() for Bayesian posterior classification. These methods predate and are more general than the binary-character similarity coefficients of Sokal and Sneath (1963) that appear in other R packages.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.1

Suggests testthat (>= 3.0.0), knitr, rmarkdown, ggplot2

VignetteBuilder knitr

URL https://github.com/AkzhanBerdi/lyubishchev-r

BugReports https://github.com/AkzhanBerdi/lyubishchev-r/issues

NeedsCompilation no

Author Akzhan Berdeyev [aut, cre]

Maintainer Akzhan Berdeyev <akzhan.berdeyev@gmail.com>

Repository https://cran.r-universe.dev

Date/Publication 2026-06-22 18:32:08 UTC

RemoteUrl https://github.com/cran/lyubishchev

RemoteRef HEAD

RemoteSha a09cfd7196d53d7b7e16d6283d1dd0bd0a92802d

Contents

`classify` *Classify a Specimen by Multivariate Posterior Probability*

Description

Assigns posterior class probabilities to a new specimen using the Edgeworth-Pearson multivariate Gaussian likelihood for each class scatter ellipse. For each class the log-likelihood of the specimen under a multivariate normal with the class mean and covariance is computed, and a softmax over the per-class log-likelihoods yields posterior probabilities.

Usage

```
classify(specimen, ellipses)
```

Arguments

`specimen`	A numeric vector of feature values for a single observation.
`ellipses`	A named list of scatter ellipses as returned by `scatter_ellipse`.

Details

The log-likelihood for class k is

$$-\frac{1}{2}\left(p\log 2\pi + \log|\Sigma_k| + (x-\mu_k)^\top\Sigma_k^{-1}(x-\mu_k)\right)$$

where p is the number of features, μ_k and Σ_k are the class mean and covariance, and x is the specimen.

Value

A named list with one element per class. Each element is a list with components:

mahalanobis_distance Squared Mahalanobis distance from the specimen to the class centroid.

log_likelihood Multivariate Gaussian log-likelihood of the specimen under the class.

posterior Posterior probability of the class (softmax over the per-class log-likelihoods). Posteriors sum to 1 across classes.

References

Lubischew, A.A. (1962). On the use of discriminant functions in taxonomy. Biometrics, 18(4), 455-477.

See Also

scatter_ellipse

Examples

```
ellipses <- scatter_ellipse(iris[, 1:4], iris$Species)
specimen <- c(5.1, 3.5, 1.4, 0.2)
result <- classify(specimen, ellipses)
sapply(result, function(r) r$posterior)
```

`divergence_coefficient`

Lyubishchev's Divergence Coefficient

Description

Computes Lyubishchev's divergence coefficient D between two groups measured on one or more continuous features. The coefficient summarises the standardised separation between the group means, summed across features:

$$D = \sum_j \frac{(M_{1j} - M_{2j})^2}{\sigma_{1j}^2 + \sigma_{2j}^2}$$

where M_{ij} and σ_{ij}^2 are the mean and (sample) variance of feature j in group i. Features whose pooled variance is zero are skipped to avoid division by zero.

Usage

```
divergence_coefficient(a, b)
```

Arguments

`a`	A numeric matrix or data frame for the first group, with one row per observation and one column per feature. A numeric vector is treated as a single-feature group.
`b`	A numeric matrix or data frame for the second group, with the same columns (features) as `a`.

Details

This is the measure described in Lyubishchev's 1943 manuscript and later published in English by Lubischew (1962). It predates and is more general than the binary-character similarity coefficients of Sokal and Sneath (1963), operating directly on continuous measurements.

Value

A single numeric value, the divergence coefficient D. Larger values indicate greater separation between the groups.

References

Lyubishchev, A.A. (1943). Programma obshchey sistematiki [Program of General Systematics]. Manuscript, 22 November 1943.

Lubischew, A.A. (1962). On the use of discriminant functions in taxonomy. Biometrics, 18(4), 455-477.

Examples

```
setosa <- as.matrix(iris[iris$Species == "setosa", 1:4])
versicolor <- as.matrix(iris[iris$Species == "versicolor", 1:4])
divergence_coefficient(setosa, versicolor)
```

`scatter_ellipse` *Fit Scatter Ellipses per Class*

Description

Fits a covariance ellipse to each class in a labelled multivariate data set. For every class the function computes the centroid (mean vector), the feature covariance matrix and the sample size. These ellipses are the building blocks for `transgression` and `classify`.

Usage

```
scatter_ellipse(X, y)
```

Arguments

`X`	A numeric matrix or data frame of observations, with one row per observation and one column per feature.
`y`	A vector of class labels of length `nrow(X)`. May be a factor, character or numeric vector.

Value

A named list with one element per class. Each element is itself a list with components:

mean Numeric vector of feature means for the class.

cov Feature covariance matrix for the class.

n_samples Integer count of observations in the class.

The names of the list are the class labels (coerced to character).

References

Lubischew, A.A. (1962). On the use of discriminant functions in taxonomy. Biometrics, 18(4), 455-477.

See Also

`transgression`, `classify`

Examples

```
ellipses <- scatter_ellipse(iris[, 1:4], iris$Species)
ellipses[["setosa"]]$mean
ellipses[["setosa"]]$n_samples
```

`transgression` *Detect Overlap (Transgression) Between Two Scatter Ellipses*

Description

Tests whether two class scatter ellipses overlap, in Lyubishchev's sense of "transgression" between groups. The centroids are compared using the squared Mahalanobis distance under the pooled covariance of the two classes, and that distance is compared against a chi-squared threshold with degrees of freedom equal to the number of features. When the Mahalanobis distance is below the threshold the groups are deemed to transgress (overlap).

Usage

```
transgression(ellipses, class_a, class_b, confidence = 0.95)
```

Arguments

`ellipses`	A named list of scatter ellipses as returned by `scatter_ellipse`.
`class_a`	Name (character) of the first class in `ellipses`.
`class_b`	Name (character) of the second class in `ellipses`.
`confidence`	Confidence level for the chi-squared threshold, between 0 and 1. Defaults to 0.95.

Value

A list with components:

mahalanobis_distance Squared Mahalanobis distance between the two centroids under the pooled covariance.

threshold Chi-squared threshold at the requested confidence with degrees of freedom equal to the number of features.

transgression Logical; `TRUE` when the distance is below the threshold (the ellipses overlap).

separation_ratio Ratio of the Mahalanobis distance to the threshold. Values above 1 indicate well-separated groups.

References

Lubischew, A.A. (1962). On the use of discriminant functions in taxonomy. Biometrics, 18(4), 455-477.

See Also

`scatter_ellipse`

Examples

```
ellipses <- scatter_ellipse(iris[, 1:4], iris$Species)
transgression(ellipses, "versicolor", "virginica")
```

Index