Package 'loclm'

August 7, 2026

Title Local Linear Regression

Version 1.0.0

Description A bare-bones implementation of local linear regression, using high-level functions. Can handle both numerical and non-numerical (factor) variables. See Loader (1999) <doi:10.1007/b98858>, ch2, or Hastie et al (2009) <doi:10.1007/978-0-387-84858-7>, ch6. The package also contains a scale_df() function which only scales the numeric variables in a dataframe.

License GPL (>= 3)

Encoding UTF-8

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports stats

Depends R (>= 4.1.0)

Config/roxygen2/version 8.0.0

VignetteBuilder knitr

NeedsCompilation no

Author Jonathan Rougier [aut, cre] (ORCID:
<https://orcid.org/0000-0003-3072-7043>)

Maintainer Jonathan Rougier <j.c.rougier@bristol.ac.uk>

Repository CRAN

Date/Publication 2026-08-07 16:30:02 UTC

Contents

hatvalues.loclm *Hat Values for Local Linear Regression*

Description

Compute the diagonal of the hat matrix. The trace of the hat matrix (the sum of the diagonal) measures the effective number of parameters.

Usage

```
## S3 method for class 'loclm'
hatvalues(model, ...)
```

Arguments

model	Object with S3 class "loclm", see loclm().
...	For compatibility with S3 generic hatvalues().

Value

A vector of hat values.

Examples

```
## see ?loclm
```

loclm *Local Linear Regression*

Description

A bare-bones implementation of local linear regression, using high-level functions. Can handle any number of regressors, both numerical and non-numerical. Non-numerical regressors are converted to factors.

Usage

```
loclm(x, y, span = 0.75)

## S3 method for class 'loclm'
predict(object, newdata, ...)

## S3 method for class 'loclm'
fitted(object, ...)

## S3 method for class 'loclm'
residuals(object, ...)
```

Arguments

x	Dataframe or numeric matrix of regressors. If a dataframe, x must contain at least one numeric variable.
y	Numeric vector of responses.
span	Proportion of points in the neighbourhood. Can be larger than one: span = Inf is equivalent to linear regression.
object	Object with S3 class "loclm".
newdata	Dataframe or matrix at which to predict, defaults to x.
...	For compatibility with S3 generic predict().

Details

Local linear regression fits a plane to the points in the neighbourhood of x, in order to predict $y(x)$. These points are weighted in the fit using the tricube weighting function. Because the neighbourhood and the weights depend on a measure of distance, numeric variables in x and newdata are prescaled using scale_df(). Factors are converted to columns of dummy variables.

The distance between different levels of the same factor (in dummy variables) is 1 or sqrt(2). The mean distance between two independent standard Normal random variables is a little over one (about 1.13). Therefore standardizing the numeric variables and dummying the factors puts all variables on roughly the same footing, as far as distance is concerned.

The behaviour with missing values can be complicated, and therefore only complete cases are used, see complete.cases().

One edge case is where a variable is a constant: scaling creates a variable of NAs. These variables are removed with a warning, and will be ignored in the prediction.

Value

A list with S3 class "loclm". This has a predict() method which returns a numeric vector of point predictions for the rows of newdata.

There are also fitted() and residuals() methods, although these are just wrappers for predict().

Quadratic regression

Modifying the code to use quadratic regression is fairly straightforward: the main difference is that the model matrix is not longer the same as the data matrix. However, there can be a lot of extra terms, especially when including factors with more than two levels. This creates a prediction with low bias but high variance, especially for extrapolation. For four or fewer numeric variables, use stats::loess() with degree = 2 (the default). Otherwise, change the code at your own risk.

References

More detail about local regression:

T. Hastie, R. Tibshirani, J. Friedman (2009), The Elements of Statistical Learning, Springer, 2nd ed, ch6.

C. Loader (1999), Local Regression and Likelihood, Springer, ch2.

See Also

stats::loess() is faster and more flexible. However, this function can only handle up to four variables, all numeric.

hatvalues.loclm() calculates hat values.

Examples

```
## loess example

fit <- loclm(cars[, "speed", drop=FALSE], cars$dist)
newd <- data.frame(speed = seq(5, 30, 1))
yhat <- predict(fit, newd)
cex <- 0.5
plot(cars$speed, residuals(fit), pch = 19, cex = cex,
  main = "Residual plot")

## stackloss data

x <- stack.x
y <- stack.loss
fit <- loclm(x, y, span = 0.5)
yhat <- fitted(fit)
plot(y, yhat, pch = 19, cex = 0.8, xlab = "Actual",
  ylab = "Predicted", main = "Stackloss dataset", asp = 1)
abline(a = 0, b = 1, lty = 2)

## add bogus factor

x <- as.data.frame(x)
x$ferret <- factor(rep(LETTERS[1:3], length.out = nrow(x)))
y <- y + as.numeric(x$ferret)

fit <- loclm(x, y, span = 0.5)
show(head(fit$x)) # now with ferret!
yhat <- fitted(fit)
plot(y, yhat, pch = 19, cex = 0.8, xlab = "Actual",
  ylab = "Predicted", main = "Stackloss dataset, with ferret",
  asp = 1)
abline(a = 0, b = 1, lty = 2)

## hat values

infl <- hatvalues(fit)
show(sum(infl)) # 17.6, effective number of parameters
fit0 <- loclm(x, y, span = Inf) # linear
infl0 <- hatvalues(fit0)
show(sum(infl0)) # 6 parameters


## some timings

big.x <- data.frame(A = x, B = x)
```

```
big.x <- do.call("rbind", rep(list(big.x), 100)) # 2100 x 8
big.y <- rep(y, length.out = nrow(big.x))
system.time({
  fit <- loclm(big.x, big.y)
  yhat <- predict(fit, big.x)
}) # 1.3s
```

`scale_df` *Scale a Dataframe*

Description

Only rescale the numeric variables in a dataframe.

Warning: does not retain the default functionality of `base::scale()`, because the returned object is always a dataframe. If `base::scale()` is called with a dataframe it will convert the returned object into a matrix.

Usage

```
scale_df(x, center = TRUE, scale = TRUE)
```

Arguments

`x`	Dataframe.
`center`	Either a logical value or a numeric vector of length equal to the number of numeric columns of `x`.
`scale`	Either a logical value or a numeric vector of length equal to the number of numeric columns of `x`.

Details

If `center` is numeric, `names(center)` identifies the numeric columns in `x`. Otherwise, if `scale` is numeric, `names(scale)` identifies the numeric columns in `x`. `center` and `scale` are passed to `base::scale()`.

Value

A dataframe like `x`, with numeric columns centered and/or scaled. The numeric centering and scalings used (if any) are returned as attributes `"scaled:center"` and `"scaled:scale"`.

See Also

`base::scale()`, which this function uses but only on a subset of the columns.

Examples

```
n <- 11
X <- data.frame(
  x = rnorm(n),
  foo = factor(sample(LETTERS[1:3], n, replace = TRUE)),
  bar = sample(LETTERS[4:5], n, replace = TRUE),
  y = rexp(n, rate = 0.2))
print(Y <- scale_df(X))
str(Y) # see attributes

## scale only

sc <- attr(Y, "scaled:scale") # has names
print(Z <- scale_df(X, center = FALSE, scale = sc))

## no numeric variables

Y <- X[, c("foo", "bar")]
Z <- scale_df(Y)
identical(Y, Z) # TRUE

## error messages

try(Y <- scale_df(X, center = c(foo = 1, bob = 2)))
try(Y <- scale_df(X, scale = c(foo = 1, bob = 2)))
```

Index