

# Package: sgd (via r-universe)

October 13, 2024

**Type** Package

**Title** Stochastic Gradient Descent for Scalable Estimation

**Version** 1.1.2

**Maintainer** Junhyung Lyle Kim <jlylekim@gmail.com>

**Description** A fast and flexible set of tools for large scale estimation. It features many stochastic gradient methods, built-in models, visualization tools, automated hyperparameter tuning, model checking, interval estimation, and convergence diagnostics.

**URL** <https://github.com/airoldilab/sgd>

**BugReports** <https://github.com/airoldilab/sgd/issues>

**License** GPL-2

**Imports** ggplot2, MASS, methods, Rcpp (>= 0.11.3), stats

**Suggests** bigmemory, glmnet, gridExtra, R.rsp, testthat

**LinkingTo** BH, bigmemory, Rcpp, RcppArmadillo

**LazyData** yes

**VignetteBuilder** R.rsp

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**Repository** <https://airoldilab.r-universe.dev>

**RemoteUrl** <https://github.com/airoldilab/sgd>

**RemoteRef** HEAD

**RemoteSha** 92246a3e1d81c9cd3077b80cacb6a3f7ee2425c5

## Contents

coef.sgd . . . . .	2
fitted.sgd . . . . .	2
plot.sgd . . . . .	3

predict.sgd . . . . .	4
print.sgd . . . . .	4
residuals.sgd . . . . .	5
sgd . . . . .	5
winequality . . . . .	9

## Index 10

---

coef.sgd	<i>Extract Model Coefficients</i>
----------	-----------------------------------

---

### Description

Extract model coefficients from sgd objects. `coefficients` is an *alias* for it.

### Usage

```
## S3 method for class 'sgd'
coef(object, ...)
```

### Arguments

object	object of class sgd.
...	some methods for this generic require additional arguments. None are used in this method.

### Value

Coefficients extracted from the model object `object`.

---

fitted.sgd	<i>Extract Model Fitted Values</i>
------------	------------------------------------

---

### Description

Extract fitted values from from sgd objects. `fitted.values` is an *alias* for it.

### Usage

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

### Arguments

object	object of class sgd.
...	some methods for this generic require additional arguments. None are used in this method.

**Value**

Fitted values extracted from the object object.

---

plot.sgd

*Plot objects of class sgd.*

---

**Description**

Plot objects of class sgd.

**Usage**

```
## S3 method for class 'sgd'
plot(x, ..., type = "mse", xaxis = "iteration")

## S3 method for class 'list'
plot(x, ..., type = "mse", xaxis = "iteration")
```

**Arguments**

x	object of class sgd.
...	additional arguments used for each type of plot. See ‘Details’.
type	character specifying the type of plot: "mse", "clf", "mse-param". See ‘Details’. Default is "mse".
xaxis	character specifying the x-axis of plot: "iteration" plots the y values over the log-iteration of the algorithm; "runtime" plots the y values over the time in seconds to reach them. Default is "iteration".

**Details**

Types of plots available:

mse Mean squared error in predictions, which takes the following arguments:

x\_test test set  
y\_test test responses to compare predictions to

clf Classification error in predictions, which takes the following arguments:

x\_test test set  
y\_test test responses to compare predictions to

mse-param Mean squared error in parameters, which takes the following arguments:

true\_param true vector of parameters to compare to

---

predict.sgd

*Model Predictions*

---

### Description

Form predictions using the estimated model parameters from stochastic gradient descent.

### Usage

```
## S3 method for class 'sgd'
predict(object, newdata, type = "link", ...)

predict_all(object, newdata, ...)
```

### Arguments

object	object of class sgd.
newdata	design matrix to form predictions on
type	the type of prediction required. The default "link" is on the scale of the linear predictors; the alternative "response" is on the scale of the response variable. Thus for a default binomial model the default predictions are of log-odds (probabilities on logit scale) and 'type = "response"' gives the predicted probabilities. The "terms" option returns a matrix giving the fitted values of each term in the model formula on the linear predictor scale.
...	further arguments passed to or from other methods.

### Details

A column of 1's must be included to newdata if the parameters include a bias (intercept) term.

---

print.sgd

*Print objects of class sgd.*

---

### Description

Print objects of class sgd.

### Usage

```
## S3 method for class 'sgd'
print(x, ...)
```

### Arguments

x	object of class sgd.
...	further arguments passed to or from other methods.

---

residuals.sgd	<i>Extract Model Residuals</i>
---------------	--------------------------------

---

### Description

Extract model residuals from sgd objects. resid is an *alias* for it.

### Usage

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

### Arguments

object	object of class sgd.
...	some methods for this generic require additional arguments. None are used in this method.

### Value

Residuals extracted from the object object.

---

sgd	<i>Stochastic gradient descent</i>
-----	------------------------------------

---

### Description

Run stochastic gradient descent in order to optimize the induced loss function given a model and data.

### Usage

```
sgd(x, ...)  
  
## S3 method for class 'formula'  
sgd(formula, data, model, model.control = list(), sgd.control = list(...), ...)  
  
## S3 method for class 'matrix'  
sgd(x, y, model, model.control = list(), sgd.control = list(...), ...)  
  
## S3 method for class 'big.matrix'  
sgd(x, y, model, model.control = list(), sgd.control = list(...), ...)
```

**Arguments**

<code>x, y</code>	a design matrix and the respective vector of outcomes.
<code>...</code>	arguments to be used to form the default <code>sgd.control</code> arguments if it is not supplied directly.
<code>formula</code>	an object of class <code>"formula"</code> (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details can be found in <code>"glm"</code> .
<code>data</code>	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>glm</code> is called.
<code>model</code>	character specifying the model to be used: <code>"lm"</code> (linear model), <code>"glm"</code> (generalized linear model), <code>"cox"</code> (Cox proportional hazards model), <code>"gmm"</code> (generalized method of moments), <code>"m"</code> (M-estimation). See 'Details'.
<code>model.control</code>	<p>a list of parameters for controlling the model.</p> <p><code>family("glm")</code> a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See <code>family</code> for details of family functions.)</p> <p><code>rank("glm")</code> logical. Should the rank of the design matrix be checked?</p> <p><code>fn("gmm")</code> a function <math>g(\theta, x)</math> which returns a <math>k</math>-vector corresponding to the <math>k</math> moment conditions. It is a required argument if <code>gr</code> not specified.</p> <p><code>gr("gmm")</code> a function to return the gradient. If unspecified, a finite-difference approximation will be used.</p> <p><code>nparams("gmm")</code> number of model parameters. This is automatically determined for other models.</p> <p><code>type("gmm")</code> character specifying the generalized method of moments procedure: <code>"twostep"</code> (Hansen, 1982), <code>"iterative"</code> (Hansen et al., 1996). Defaults to <code>"iterative"</code>.</p> <p><code>wmatrix("gmm")</code> weighting matrix to be used in the loss function. Defaults to the identity matrix.</p> <p><code>loss("m")</code> character specifying the loss function to be used in the estimating equation. Default is the Huber loss.</p> <p><code>lambda1</code> L1 regularization parameter. Default is 0.</p> <p><code>lambda2</code> L2 regularization parameter. Default is 0.</p>
<code>sgd.control</code>	<p>an optional list of parameters for controlling the estimation.</p> <p><code>method</code> character specifying the method to be used: <code>"sgd"</code>, <code>"implicit"</code>, <code>"asgd"</code>, <code>"ai-sgd"</code>, <code>"momentum"</code>, <code>"nesterov"</code>. Default is <code>"ai-sgd"</code>. See 'Details'.</p> <p><code>lr</code> character specifying the learning rate to be used: <code>"one-dim"</code>, <code>"one-dim-eigen"</code>, <code>"d-dim"</code>, <code>"adagrad"</code>, <code>"rmsprop"</code>. Default is <code>"one-dim"</code>. See 'Details'.</p> <p><code>lr.control</code> vector of scalar hyperparameters one can set dependent on the learning rate. For hyperparameters aimed to be left as default, specify NA in the corresponding entries. See 'Details'.</p> <p><code>start</code> starting values for the parameter estimates. Default is random initialization around zero.</p>

size number of SGD estimates to store for diagnostic purposes (distributed log-uniformly over total number of iterations)

reltol relative convergence tolerance. The algorithm stops if it is unable to change the relative mean squared difference in the parameters by more than the amount. Default is  $1e-05$ .

npasses the maximum number of passes over the data. Default is 3.

pass logical. Should tol be ignored and run the algorithm for all of npasses?

shuffle logical. Should the algorithm shuffle the data set including for each pass?

verbose logical. Should the algorithm print progress?

## Details

Models: The Cox model assumes that the survival data is ordered when passed in, i.e., such that the risk set of an observation  $i$  is all data points after it.

Methods:

sgd stochastic gradient descent (Robbins and Monro, 1951)

implicit implicit stochastic gradient descent (Toulis et al., 2014)

asgd stochastic gradient with averaging (Polyak and Juditsky, 1992)

ai-sgd implicit stochastic gradient with averaging (Toulis et al., 2015)

momentum "classical" momentum (Polyak, 1964)

nesterov Nesterov's accelerated gradient (Nesterov, 1983)

Learning rates and hyperparameters:

one-dim scalar value prescribed in Xu (2011) as

$$a_n = scale * gamma / (1 + alpha * gamma * n)^{(c)}$$

where the defaults are `lr.control = (scale=1, gamma=1, alpha=1, c)` where  $c$  is 1 if implemented without averaging, 2/3 if with averaging

one-dim-eigen diagonal matrix `lr.control = NULL`

d-dim diagonal matrix `lr.control = (epsilon=1e-6)`

adagrad diagonal matrix prescribed in Duchi et al. (2011) as `lr.control = (eta=1, epsilon=1e-6)`

rmsprop diagonal matrix prescribed in Tieleman and Hinton (2012) as `lr.control = (eta=1, gamma=0.9, epsilon=1e-6)`

## Value

An object of class "sgd", which is a list containing the following components:

model	name of the model
coefficients	a named vector of coefficients
converged	logical. Was the algorithm judged to have converged?
estimates	estimates from algorithm stored at each iteration specified in pos

fitted.values	the fitted mean values
pos	vector of indices specifying the iteration number each estimate was stored for
residuals	the residuals, that is response minus fitted values
times	vector of times in seconds it took to complete the number of iterations specified in pos
model.out	a list of model-specific output attributes

### Author(s)

Dustin Tran, Tian Lan, Panos Toulis, Ye Kuang, Edoardo Airoldi

### References

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121-2159, 2011.

Yurii Nesterov. A method for solving a convex programming problem with convergence rate  $O(1/k^2)$ . *Soviet Mathematics Doklady*, 27(2):372-376, 1983.

Boris T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1-17, 1964.

Boris T. Polyak and Anatoli B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838-855, 1992.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pp. 400-407, 1951.

Panos Toulis, Jason Rennie, and Edoardo M. Airoldi, "Statistical analysis of stochastic gradient methods for generalized linear models", In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

Panos Toulis, Dustin Tran, and Edoardo M. Airoldi, "Stability and optimality in stochastic gradient descent", arXiv preprint arXiv:1505.02417, 2015.

Wei Xu. Towards optimal one pass large scale learning with averaged stochastic gradient descent. arXiv preprint arXiv:1107.2490, 2011.

# Dimensions

### Examples

```
## Linear regression
set.seed(42)
N <- 1e4
d <- 5
X <- matrix(rnorm(N*d), ncol=d)
theta <- rep(5, d+1)
eps <- rnorm(N)
y <- cbind(1, X) %*% theta + eps
dat <- data.frame(y=y, x=X)
sgd.theta <- sgd(y ~ ., data=dat, model="lm")
sprintf("Mean squared error: %0.3f", mean((theta - as.numeric(sgd.theta$coefficients))^2))
```



---

`winequality`*Wine quality data of white wine samples from Portugal*

---

**Description**

This dataset is a collection of white "Vinho Verde" wine samples from the north of Portugal. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

**Usage**`winequality`**Format**

A data frame with 4898 rows and 12 variables

- fixed acidity.
- volatile acidity.
- citric acid.
- residual sugar.
- chlorides.
- free sulfur dioxide.
- total sulfur dioxide.
- density.
- pH.
- sulphates.
- alcohol.
- quality (score between 0 and 10).

**Source**

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

# Index

## \* datasets

winequality, 9

as.data.frame, 6

coef.sgd, 2

family, 6

fitted.sgd, 2

formula, 6

glm, 6

plot.list (plot.sgd), 3

plot.sgd, 3

predict.sgd, 4

predict\_all (predict.sgd), 4

print.sgd, 4

residuals.sgd, 5

sgd, 5

winequality, 9