

parenttoc: True

## Generated datasets

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 10)**

Unknown directive type "currentmodule".

```
.. currentmodule:: sklearn.datasets
```

In addition, scikit-learn includes various random sample generators that can be used to build artificial datasets of controlled size and complexity.

### Generators for classification and clustering

These generators produce a matrix of features and corresponding discrete targets.

#### Single label

Both `:func:`make_blobs`` and `:func:`make_classification`` create multiclass datasets by allocating each class one or more normally-distributed clusters of points. `:func:`make_blobs`` provides greater control regarding the centers and standard deviations of each cluster, and is used to demonstrate clustering. `:func:`make_classification`` specialises in introducing noise by way of: correlated, redundant and uninformative features; multiple Gaussian clusters per class; and linear transformations of the feature space.

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 24);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 24);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 24);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 24);**  
[backlink](#)

Unknown interpreted text role "func".

`:func:`make_gaussian_quantiles`` divides a single Gaussian cluster into near-equal-size classes separated by concentric hyperspheres. `:func:`make_hastie_10_2`` generates a similar binary, 10-dimensional problem.

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 32);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 32);**  
[backlink](#)

Unknown interpreted text role "func".

`:func:'make_circles'` and `:func:'make_moons'` generate 2d binary classification datasets that are challenging to certain algorithms (e.g. centroid-based clustering or linear classification), including optional Gaussian noise. They are useful for visualisation. `:func:'make_circles'` produces Gaussian data with a spherical decision boundary for binary classification, while `:func:'make_moons'` produces two interleaving half circles.

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 41);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 41);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 41);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 41);**  
[backlink](#)

Unknown interpreted text role "func".

## Multilabel

`:func:'make_multilabel_classification'` generates random samples with multiple labels, reflecting a bag of words drawn from a mixture of topics. The number of topics for each document is drawn from a Poisson distribution, and the topics themselves are drawn from a fixed random distribution. Similarly, the number of words is drawn from Poisson, with words drawn from a multinomial, where each topic defines a probability distribution over words. Simplifications with respect to true bag-of-words mixtures include:

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 51);**  
[backlink](#)

Unknown interpreted text role "func".

- Per-topic word distributions are independently drawn, where in reality all would be affected by a sparse base distribution, and would be correlated.
- For a document generated from multiple topics, all topics are weighted equally in generating its bag of words.
- Documents without labels words at random, rather than from a base distribution.

## Biclustering

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 74)**

Unknown directive type "autosummary".

```
.. autosummary::  
    make_biclusters  
    make_checkerboard
```

## Generators for regression

`:func:'make_regression'` produces regression targets as an optionally-sparse random linear combination of random features, with noise. Its informative features may be uncorrelated, or low rank (few features account for most of the variance).

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 83);**  
[backlink](#)

Unknown interpreted text role "func".

Other regression generators generate functions deterministically from randomized features. `:func:`make_sparse_unrelated`` produces a target as a linear combination of four features with fixed coefficients. Others encode explicitly non-linear relations: `:func:`make_friedman1`` is related by polynomial and sine transforms; `:func:`make_friedman2`` includes feature multiplication and reciprocation; and `:func:`make_friedman3`` is similar with an arctan transformation on the target.

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 88);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 88);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 88);**  
[backlink](#)

Unknown interpreted text role "func".

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 88);**  
[backlink](#)

Unknown interpreted text role "func".

## Generators for manifold learning

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 99)**

Unknown directive type "autosummary".

```
.. autosummary::  
  
    make_s_curve  
    make_swiss_roll
```

## Generators for decomposition

**System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\datasets\[scikit-learn-main] [doc] [datasets] sample\_generators.rst, line 107)**

Unknown directive type "autosummary".

```
.. autosummary::  
  
    make_low_rank_matrix  
    make_sparse_coded_signal  
    make_spd_matrix  
    make_sparse_spd_matrix
```