

Utilities for Developers

Scikit-learn contains a number of utilities to help with development. These are located in `:mod:'sklearn.utils'`, and include tools in a number of categories. All the following functions and classes are in the module `:mod:'sklearn.utils'`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 7); [backlink](#)

Unknown interpreted text role "mod".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 7); [backlink](#)

Unknown interpreted text role "mod".

Warning

These utilities are meant to be used internally within the scikit-learn package. They are not guaranteed to be stable between versions of scikit-learn. Backports, in particular, will be removed as the scikit-learn dependencies evolve.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 19)

Unknown directive type "currentmodule".

```
.. currentmodule:: sklearn.utils
```

Validation Tools

These are tools used to check and validate input. When you write a function which accepts arrays, matrices, or sparse matrices as arguments, the following should be used when applicable.

- `:func:'assert_all_finite'`: Throw an error if array contains NaNs or Infs.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 28); [backlink](#)

Unknown interpreted text role "func".

- `:func:'as_float_array'`: convert input to an array of floats. If a sparse matrix is passed, a sparse matrix will be returned.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 30); [backlink](#)

Unknown interpreted text role "func".

- `:func:'check_array'`: check that input is a 2D array, raise error on sparse matrices. Allowed sparse matrix formats can be given optionally, as well as allowing 1D or N-dimensional arrays. Calls `:func:'assert_all_finite'` by default.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 33); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 33); [backlink](#)

Unknown interpreted text role "func".

- `:func:`check_X_y``: check that X and y have consistent length, calls `check_array` on X, and `column_or_1d` on y. For multilabel classification or multitarget regression, specify `multi_output=True`, in which case `check_array` will be called on y.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 38); [backlink](#)

Unknown interpreted text role "func".

- `:func:`indexable``: check that all input arrays have consistent length and can be sliced or indexed using `safe_index`. This is used to validate input for cross-validation.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 43); [backlink](#)

Unknown interpreted text role "func".

- `:func:`validation.check_memory`` checks that input is `joblib.Memory`-like, which means that it can be converted into a `sklearn.utils.Memory` instance (typically a str denoting the `cachedir`) or has the same interface.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 47); [backlink](#)

Unknown interpreted text role "func".

If your code relies on a random number generator, it should never use functions like `numpy.random.random` or `numpy.random.normal`. This approach can lead to repeatability issues in unit tests. Instead, a `numpy.random.RandomState` object should be used, which is built from a `random_state` argument passed to the class or function. The function `:func:`check_random_state``, below, can then be used to create a random number generator object.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 52); [backlink](#)

Unknown interpreted text role "func".

- `:func:`check_random_state``: create a `np.random.RandomState` object from a parameter `random_state`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 60); [backlink](#)

Unknown interpreted text role "func".

- If `random_state` is `None` or `np.random`, then a randomly-initialized `RandomState` object is returned.
- If `random_state` is an integer, then it is used to seed a new `RandomState` object.
- If `random_state` is a `RandomState` object, then it is passed through.

For example:

```
>>> from sklearn.utils import check_random_state
>>> random_state = 0
>>> random_state = check_random_state(random_state)
>>> random_state.rand(4)
array([0.5488135 , 0.71518937, 0.60276338, 0.54488318])
```

When developing your own scikit-learn compatible estimator, the following helpers are available.

- `:func:`validation.check_is_fitted``: check that the estimator has been fitted before calling `transform`, `predict`, or similar methods. This helper allows to raise a standardized error message across estimator.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 80); [backlink](#)

Unknown interpreted text role "func".

- `:func:`validation.has_fit_parameter``: check that a given parameter is supported in the `fit` method of a given estimator.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 84); [backlink](#)

Unknown interpreted text role "func".

Efficient Linear Algebra & Array Operations

- `:func:`extmath.randomized_range_finder``: construct an orthonormal matrix whose range approximates the range of the input. This is used in `:func:`extmath.randomized_svd``, below.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 90); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 90); [backlink](#)

Unknown interpreted text role "func".

- `:func:`extmath.randomized_svd``: compute the k-truncated randomized SVD. This algorithm finds the exact truncated singular values decomposition using randomization to speed up the computations. It is particularly fast on large matrices on which you wish to extract only a small number of components.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 94); [backlink](#)

Unknown interpreted text role "func".

- `:func:`arrayfuncs.cholesky_delete`` (used in `:func:`~sklearn.linear_model.lars_path``) Remove an item from a cholesky factorization.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 100); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 100); [backlink](#)

Unknown interpreted text role "func".

- `:func:`arrayfuncs.min_pos`` (used in `sklearn.linear_model.least_angle`) Find the minimum of the positive values within an array.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 104); [backlink](#)

Unknown interpreted text role "func".

- `:func:`extmath.fast_logdet``: efficiently compute the log of the determinant of a matrix.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-

resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 108); [backlink](#)

Unknown interpreted text role "func".

- `:func:`extmath.density``: efficiently compute the density of a sparse vector

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 111); [backlink](#)

Unknown interpreted text role "func".

- `:func:`extmath.safe_sparse_dot``: dot product which will correctly handle `scipy.sparse` inputs. If the inputs are dense, it is equivalent to `numpy.dot`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 113); [backlink](#)

Unknown interpreted text role "func".

- `:func:`extmath.weighted_mode``: an extension of `scipy.stats.mode` which allows each item to have a real-valued weight.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 117); [backlink](#)

Unknown interpreted text role "func".

- `:func:`resample``: Resample arrays or sparse matrices in a consistent way. used in `:func:`shuffle``, below.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 120); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 120); [backlink](#)

Unknown interpreted text role "func".

- `:func:`shuffle``: Shuffle arrays or sparse matrices in a consistent way. Used in `:func:`~sklearn.cluster.k_means``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 123); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]
[developers]utilities.rst, line 123); [backlink](#)

Unknown interpreted text role "func".

Efficient Random Sampling

- `:func:`random.sample_without_replacement``: implements efficient algorithms for sampling `n_samples` integers from a population of size `n_population` without replacement.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc]

[developers]utilities.rst, line 130); [backlink](#)

Unknown interpreted text role "func".

Efficient Routines for Sparse Matrices

The `sklearn.utils.sparsefuncs` cython module hosts compiled extensions to efficiently process `scipy.sparse` data.

- `:func:`sparsefuncs.mean_variance_axis``: compute the means and variances along a specified axis of a CSR matrix. Used for normalizing the tolerance stopping criterion in `:class:`~sklearn.cluster.KMeans``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 141); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 141); [backlink](#)

Unknown interpreted text role "class".

- `:func:`sparsefuncs_fast.inplace_csr_row_normalize_l1`` and `:func:`sparsefuncs_fast.inplace_csr_row_normalize_l2``: can be used to normalize individual sparse samples to unit L1 or L2 norm as done in `:class:`~sklearn.preprocessing.Normalizer``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 146); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 146); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 146); [backlink](#)

Unknown interpreted text role "class".

- `:func:`sparsefuncs.inplace_csr_column_scale``: can be used to multiply the columns of a CSR matrix by a constant scale (one scale per column). Used for scaling features to unit standard deviation in `:class:`~sklearn.preprocessing.StandardScaler``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 151); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 151); [backlink](#)

Unknown interpreted text role "class".

Graph Routines

- `:func:`graph.single_source_shortest_path_length``: (not currently used in scikit-learn) Return the shortest path from a single source to all connected nodes on a graph. Code is adapted from [networkx](#). If this is ever needed again, it would be far faster to use a single iteration of Dijkstra's algorithm from `graph_shortest_path`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 160); [backlink](#)

Unknown interpreted text role "func".

Testing Functions

- `:func:`all_estimators``: returns a list of all estimators in scikit-learn to test for consistent behavior and interfaces.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 172); [backlink](#)

Unknown interpreted text role "func".

Multiclass and multilabel utility function

- `:func:`multiclass.is_multilabel``: Helper function to check if the task is a multi-label classification one.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 178); [backlink](#)

Unknown interpreted text role "func".

- `:func:`multiclass.unique_labels``: Helper function to extract an ordered array of unique labels from different formats of target.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 181); [backlink](#)

Unknown interpreted text role "func".

Helper Functions

- `:class:`gen_even_slices``: generator to create n -packs of slices going up to n . Used in `:func:`~sklearn.decomposition.dict_learning`` and `:func:`~sklearn.cluster.k_means``.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 188); [backlink](#)

Unknown interpreted text role "class".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 188); [backlink](#)

Unknown interpreted text role "func".

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 188); [backlink](#)

Unknown interpreted text role "func".

- `:class:`gen_batches``: generator to create slices containing batch size elements from 0 to n

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main] [doc] [developers]utilities.rst, line 192); [backlink](#)

Unknown interpreted text role "class".

- `:func:'safe_mask'`: Helper function to convert a mask to the format expected by the numpy array or scipy sparse matrix on which to use it (sparse matrices support integer indices only while numpy arrays support both boolean masks and integer indices).

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 195); [backlink](#)

Unknown interpreted text role "func".

- `:func:'safe_sqr'`: Helper function for unified squaring ($**2$) of array-likes, matrices and sparse matrices.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 200); [backlink](#)

Unknown interpreted text role "func".

Hash Functions

- `:func:'murmurhash3_32'` provides a python wrapper for the `MurmurHash3_x86_32` C++ non cryptographic hash function. This hash function is suitable for implementing lookup tables, Bloom filters, Count Min Sketch, feature hashing and implicitly defined sparse random projections:

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 207); [backlink](#)

Unknown interpreted text role "func".

```
>>> from sklearn.utils import murmurhash3_32
>>> murmurhash3_32("some feature", seed=0) == -384616559
True

>>> murmurhash3_32("some feature", seed=0, positive=True) == 3910350737
True
```

The `sklearn.utils.murmurhash` module can also be "cimported" from other cython modules so as to benefit from the high performance of `MurmurHash` while skipping the overhead of the Python interpreter.

Warnings and Exceptions

- `:class:'deprecated'`: Decorator to mark a function or class as deprecated.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 228); [backlink](#)

Unknown interpreted text role "class".

- `:class:'~sklearn.exceptions.ConvergenceWarning'`: Custom warning to catch convergence problems. Used in `sklearn.covariance.graphical_lasso`.

System Message: ERROR/3 (D:\onboarding-resources\sample-onboarding-resources\scikit-learn-main\doc\developers\[scikit-learn-main][doc][developers]utilities.rst, line 230); [backlink](#)

Unknown interpreted text role "class".