[sklearn.datasets](#).fetch_california_housing

sklearn.datasets.**fetch_california_housing**(*, *data_home=None*, *download_if_missing=True*, *return_X_y=False*, *as_frame=False*)

[source]

Load the California housing dataset (regression).

Samples total	20640
Dimensionality	8
Features	real
Target	real 0.15 - 5.

◀ ▶

Read more in the [User Guide](#).

Parameters:

data_home : *str or path-like, default=None*

Specify another download and cache folder for the datasets. By default all scikit-learn data is stored in '~/scikit_learn_data' subfolders.

download_if_missing : *bool, default=True*

If False, raise an OSError if the data is not locally available instead of trying to download the data from the source site.

return_X_y : *bool, default=False*

If True, returns `(data.data, data.target)` instead of a Bunch object.

New in version 0.20.

as_frame : *bool, default=False*

If True, the data is a pandas DataFrame including columns with appropriate dtypes (numeric, string or categorical). The target is a pandas DataFrame or Series depending on the number of target_columns.

New in version 0.23.

Returns:

dataset : *[Bunch](#)*

Dictionary-like object, with the following attributes.

data : *ndarray, shape (20640, 8)*

Each row corresponding to the 8 feature values in order. If `as_frame` is True, `data` is a pandas object.

target : *numpy array of shape (20640,)*

Each value corresponds to the average house value in units of 100,000. If `as_frame` is True, `target` is a pandas object.

feature_names : *list of length 8*

Array of ordered feature names used in the dataset.

DESCR : *str*

Description of the California housing dataset.

frame : *pandas DataFrame*

Only present when `as_frame=True`. DataFrame with `data` and `target`.

New in version 0.23.

(data, target) : *tuple if `return_X_y` is True*

A tuple of two ndarray. The first containing a 2D array of shape (n_samples, n_features) with each row representing one sample and each column representing the features. The second ndarray of shape (n_samples,) containing the target samples.
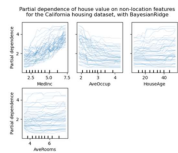
New in version 0.20.

◀ ▶

Notes

This dataset consists of 20,640 samples and 9 features.
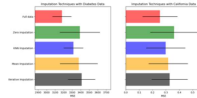
Toggle Menu

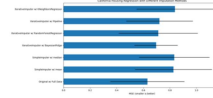# Examples using `sklearn.datasets.fetch_california_housing`

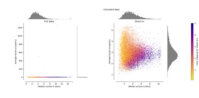Release Highlights for scikit-learn 0.24

Comparing Random Forests and Histogram Gradient Boosting models

Imputing missing values before building an estimator

Imputing missing values with variants of IterativeImputer

Compare the effect of different scalers on data with outliers

Toggle Menu