Part I

Python

Part II

Scikit-Learn

Chapter 1

feature_extraction

1.1 DictVectorizer

1.2 text

1.2.1 CounterVector

1.2.2 TfidfVectorizer

Table 1.1: feature_extraction

CounterVector	DictVectorizer	TfidfVectorizer

Chapter 2

preprocessing

2.1 PolynomialFeatures

Table 2.1: preprocessing

PolynomialFeatures

Chapter 3

impute

3.1 SimpleImputer

Table 3.1: impute

SimpleImputer

Chapter 4

pipeline

4.1 make_pipeline

Table 4.1: pipeline

make_pipeline

Chapter 5

datasets

5.1 make_blobs

5.2 make_friedman1

`make_friedman1(n_samples=100, n_features=10, *, noise=0.0, random_state=None)`

Inputs `X` are independent features uniformly distributed on the interval $[0,1]$. The output `y` is created according to the formula:

$$y = 10\sin(\pi x_1 x_2) + 20(x_3 - \frac{1}{2})^2 + 10x_4 + 5x_5 + Gaussian\ Noise(0,\sigma) \tag{5.1}$$

A synthetic data set called *Friedman-1*, originally created by Jerome Friedman in 1991 to explore how well his new multivariate adaptive regression splines (MARS) algorithm was fitting high-dimensional data.

This data set was carefully generated to evaluate a regression method's ability to only pick up true feature dependencies in the data set and ignore others.

Table 5.2: make_friedman1主要参数

Properties	Names	Descriptions
Parameters	`n_samples: int, default=100`	The number of samples.
Parameters	`n_features: int, default=10`	The number of features. Should be at least 5.
Parameters	`noise: float, default=0.0`	The standard deviation of the gaussian noise applied to the output.
Returns	`X: ndarray of shape (n_samples, n_features)`	The input samples.
Returns	`y: ndarray of shape (n_samples,)`	The output values.

Table 5.1: datasets

fetch_20newsgroups	fetch_lfw_people	make_friedman1	make_friedman2	make_friedman3	make_circles
make_blobs					

5.3 make_friedman2

5.4 make_friedman3

5.5 make_circles

5.6 fetch_20newsgroups

5.7 fetch_lfw_people

Chapter 6

naive_bayes

6.1 GaussianNB

6.2 MultinomialNB

Table 6.1: naive_bayes

GaussianNB	MultinomialNB

Chapter 7

metrics

7.1 confusion_matrix

7.2 pairwise_distances

7.3 pairwise_distances_argmin

7.4 pairwise_distances_argmax

Table 7.1: metrics

confusion_matrix	pairwise_distances	pairwise_distances_argmin	pairwise_distances_argmax

Chapter 8

linear_model

8.1 LinearRegression

8.2 Ridge

8.3 Lasso

Table 8.1: linear_model

LinearRegression	Ridge	Lasso

Chapter 9

utils

9.1 resample

Table 9.1: utils

resample

Chapter 10

svm

10.1 svc

Table 10.1: svm

svc

Chapter 11

cluster

11.1 KMean

11.2 SpectralClustering

11.3 MiniBatchKMeans

Table 11.1: cluster

KMean	SpectralClustering	MiniBatchKMeans

Part III

NumPy

Chapter 12

routines

12.1 Mathematical functions

12.1.1 prod

12.2 Set routines

12.2.1 setdiff1d

```
numpy.setdiff1d(ar1, ar2, assume_unique=False)
```

Find the set difference of two arrays.

Return the unique values in *ar1* that are not in *ar2*.

Table 12.1: routines: Mathematical functions

prod	setdiff1d