

## Elegir el número de características

Cuando se utilizan valoraciones individuales de las características puede resultar difícil elegir con cuantas características quedarse en la selección.

La clase `SelectFpr` ([https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_selection.SelectFpr.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectFpr.html)) puede ayudarnos basándose en el porcentaje de falsos positivos permitidos. El típico valor alfa de los test estadísticos. Sobre los ejemplos anteriores:

```
In [27]: import numpy as np
import pandas as pd
iris = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data",
                  names=['sepal length', 'sepal width', 'petal length', 'petal width', 'target'])
caracteristicas = ['sepal length', 'sepal width', 'petal length', 'petal width']
iris
```

Out[27]:

	sepal length	sepal width	petal length	petal width	target
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 5 columns

```
In [28]: from sklearn.feature_selection import SelectFpr
         from sklearn.feature_selection import chi2
         X = iris[caracteristicas]
         y = iris['target']

         X_new = SelectFpr(chi2, alpha=0.1).fit_transform(X, y)
         X_new
```

```
Out[28]: array([[5.1, 1.4, 0.2],
 [4.9, 1.4, 0.2],
 [4.7, 1.3, 0.2],
 [4.6, 1.5, 0.2],
 [5. , 1.4, 0.2],
 [5.4, 1.7, 0.4],
 [4.6, 1.4, 0.3],
 [5. , 1.5, 0.2],
 [4.4, 1.4, 0.2],
 [4.9, 1.5, 0.1],
 [5.4, 1.5, 0.2],
 [4.8, 1.6, 0.2],
 [4.8, 1.4, 0.1],
 [4.3, 1.1, 0.1],
 [5.8, 1.2, 0.2],
 [5.7, 1.5, 0.4],
 [5.4, 1.3, 0.4],
 [5.1, 1.4, 0.3],
 [5.7, 1.7, 0.3],
 [5.1, 1.5, 0.3],
 [5.4, 1.7, 0.2],
 [5.1, 1.5, 0.4],
 [4.6, 1. , 0.2],
 [5.1, 1.7, 0.5],
 [4.8, 1.9, 0.2],
 [5. , 1.6, 0.2],
 [5. , 1.6, 0.4],
 [5.2, 1.5, 0.2],
 [5.2, 1.4, 0.2],
 [4.7, 1.6, 0.2],
 [4.8, 1.6, 0.2],
 [5.4, 1.5, 0.4],
 [5.2, 1.5, 0.1],
 [5.5, 1.4, 0.2],
 [4.9, 1.5, 0.1],
 [5. , 1.2, 0.2],
 [5.5, 1.3, 0.2],
 [4.9, 1.5, 0.1],
 [4.4, 1.3, 0.2],
 [5.1, 1.5, 0.2],
 [5. , 1.3, 0.3],
 [4.5, 1.3, 0.3],
 [4.4, 1.3, 0.2],
 [5. , 1.6, 0.6],
 [5.1, 1.9, 0.4],
 [4.8, 1.4, 0.3],
 [5.1, 1.6, 0.2],
 [4.6, 1.4, 0.2],
 [5.3, 1.5, 0.2],
 [5. , 1.4, 0.2],
 [7. , 4.7, 1.4],
 [6.4, 4.5, 1.5],
 [6.9, 4.9, 1.5],
 [5.5, 4. , 1.3],
 [6.5, 4.6, 1.5],
 [5.7, 4.5, 1.3],
 [6.3, 4.7, 1.6],
 [4.9, 3.3, 1. ],
 [6.6, 4.6, 1.3],
 [5.2, 3.9, 1.4],
 [5. , 3.5, 1. ],
 [5.9, 4.2, 1.5],
 [6. , 4. , 1. ],
 [6.1, 4.7, 1.4],
 [5.6, 3.6, 1.3],
 [6.7, 4.4, 1.4],
 [5.6, 4.5, 1.5],
 [5.8, 4.1, 1. ],
 [6.2, 4.5, 1.5],
 [5.6, 3.9, 1.1],
 [5.9, 4.8, 1.8],
 [6.1, 4. , 1.3],
 [6.3, 4.9, 1.5],
 [6.1, 4.7, 1.2],
 [6.4, 4.3, 1.3],
 [6.6, 4.4, 1.4],
 [6.8, 4.8, 1.4],
 [6.7, 5. , 1.7],
 [6. , 4.5, 1.5],
 [5.7, 3.5, 1. ]],
```

[5.5, 3.8, 1.1],  
[5.5, 3.7, 1. ],  
[5.8, 3.9, 1.2],  
[6. , 5.1, 1.6],  
[5.4, 4.5, 1.5],  
[6. , 4.5, 1.6],  
[6.7, 4.7, 1.5],  
[6.3, 4.4, 1.3],  
[5.6, 4.1, 1.3],  
[5.5, 4. , 1.3],  
[5.5, 4.4, 1.2],  
[6.1, 4.6, 1.4],  
[5.8, 4. , 1.2],  
[5. , 3.3, 1. ],  
[5.6, 4.2, 1.3],  
[5.7, 4.2, 1.2],  
[5.7, 4.2, 1.3],  
[6.2, 4.3, 1.3],  
[5.1, 3. , 1.1],  
[5.7, 4.1, 1.3],  
[6.3, 6. , 2.5],  
[5.8, 5.1, 1.9],  
[7.1, 5.9, 2.1],  
[6.3, 5.6, 1.8],  
[6.5, 5.8, 2.2],  
[7.6, 6.6, 2.1],  
[4.9, 4.5, 1.7],  
[7.3, 6.3, 1.8],  
[6.7, 5.8, 1.8],  
[7.2, 6.1, 2.5],  
[6.5, 5.1, 2. ],  
[6.4, 5.3, 1.9],  
[6.8, 5.5, 2.1],  
[5.7, 5. , 2. ],  
[5.8, 5.1, 2.4],  
[6.4, 5.3, 2.3],  
[6.5, 5.5, 1.8],  
[7.7, 6.7, 2.2],  
[7.7, 6.9, 2.3],  
[6. , 5. , 1.5],  
[6.9, 5.7, 2.3],  
[5.6, 4.9, 2. ],  
[7.7, 6.7, 2. ],  
[6.3, 4.9, 1.8],  
[6.7, 5.7, 2.1],  
[7.2, 6. , 1.8],  
[6.2, 4.8, 1.8],  
[6.1, 4.9, 1.8],  
[6.4, 5.6, 2.1],  
[7.2, 5.8, 1.6],  
[7.4, 6.1, 1.9],  
[7.9, 6.4, 2. ],  
[6.4, 5.6, 2.2],  
[6.3, 5.1, 1.5],  
[6.1, 5.6, 1.4],  
[7.7, 6.1, 2.3],  
[6.3, 5.6, 2.4],  
[6.4, 5.5, 1.8],  
[6. , 4.8, 1.8],  
[6.9, 5.4, 2.1],  
[6.7, 5.6, 2.4],  
[6.9, 5.1, 2.3],  
[5.8, 5.1, 1.9],  
[6.8, 5.9, 2.3],  
[6.7, 5.7, 2.5],  
[6.7, 5.2, 2.3],  
[6.3, 5. , 1.9],  
[6.5, 5.2, 2. ],  
[6.2, 5.4, 2.3],  
[5.9, 5.1, 1.8]]

```
In [29]: # Cargar el conjunto de datos
from sklearn import datasets
dataset = datasets.fetch_openml(name='delta_elevators', version=1, as_frame=True)
delta = dataset.frame
delta
```

Out[29]:

	climbRate	Altitude	RollRate	curRoll	diffClb	diffDiffClb	Se
0	2.0	-50.0	-0.0048	-0.001	0.2	0.00	-0.001
1	6.5	-40.0	-0.0010	-0.009	0.2	0.00	0.003
2	-5.9	-10.0	-0.0033	-0.004	-0.1	0.00	-0.001
3	-6.2	-30.0	-0.0022	-0.011	0.1	0.00	-0.002
4	-0.2	-40.0	0.0059	-0.005	0.1	0.00	0.001
...	...	...	...	...	...	...	...
9512	5.0	-30.0	0.0013	-0.004	0.2	0.00	0.004
9513	1.4	0.0	0.0024	0.019	-0.2	-0.01	-0.001
9514	-3.5	-10.0	-0.0082	0.004	-0.1	0.00	-0.003
9515	-2.4	-10.0	-0.0065	-0.012	0.2	-0.02	-0.001
9516	4.7	-10.0	0.0018	-0.020	0.3	0.00	0.001

9517 rows × 7 columns

```
In [30]: from sklearn.feature_selection import f_regression
X = delta[delta.columns[:-1]]
y = delta['Se']
X_new = SelectFpr(f_regression, alpha=0.01).fit_transform(X, y)
X_new
```

```
Out[30]: array([[ 2.0e+00, -5.0e+01, -4.8e-03, -1.0e-03,  2.0e-01,  0.0e+00],
 [ 6.5e+00, -4.0e+01, -1.0e-03, -9.0e-03,  2.0e-01,  0.0e+00],
 [-5.9e+00, -1.0e+01, -3.3e-03, -4.0e-03, -1.0e-01,  0.0e+00],
 ...,
 [-3.5e+00, -1.0e+01, -8.2e-03,  4.0e-03, -1.0e-01,  0.0e+00],
 [-2.4e+00, -1.0e+01, -6.5e-03, -1.2e-02,  2.0e-01, -2.0e-02],
 [ 4.7e+00, -1.0e+01,  1.8e-03, -2.0e-02,  3.0e-01,  0.0e+00]])
```

**Ejercicio:** aplica selección de características basada en evaluaciones individuales a un conjunto de datos de tu elección.