



Reconocimiento de patrones

Tarea 1

Introducción: Bases matemáticas

Alumno:

Pérez Rodríguez Raúl Francisco

Octubre 2017

Rendimiento de combustible

Los primeros tres datos del conjunto

| mpg | cylinders | displacement | horsepower | weight | acceleration |
|------|-----------|--------------|------------|--------|--------------|
| 18.0 | 8 | 307.0 | 130.0 | 3504 | 12.0 |
| 15.0 | 8 | 350.0 | 165.0 | 3693 | 11.5 |
| 18.0 | 8 | 318.0 | 150.0 | 3436 | 11.0 |

| model-year | origin | car-name |
|------------|--------|---------------------------|
| 70 | 1 | chevrolet chevelle malibu |
| 70 | 1 | buick skylark 320 |
| 70 | 1 | plymouth satellite |

Informacion del data frame

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangelIndex: 398 entries, 0 to 397
```

```
Data columns (total 9 columns):
```

```
mpg          398 non-null float64
```

```
cylinders    398 non-null int64
```

```
displacement 398 non-null float64
```

```
horsepower   392 non-null float64
```

```
weight       398 non-null int64
```

```
acceleration 398 non-null float64
```

```
model-year   398 non-null int64
```

```
origin       398 non-null int64
```

```
car-name     398 non-null object
```

```
dtypes: float64(4), int64(4), object(1)
```

```
memory usage: 28.1+ KB
```

```
None
```

El conjunto de datos contiene 398 instancias. 8 de las variables son numéricas y solo la variable “car-name” es string.

| type | mpg | cylinders | displacement | horsepower | weight |
|--------------|------------|------------------|---------------------|-------------------|---------------|
| count | 398.000000 | 398.000000 | 398.000000 | 398.000000 | 398.000000 |
| mean | 23.514573 | 5.454774 | 193.425879 | 104.469388 | 2970.424623 |
| std | 7.815984 | 1.701004 | 104.269838 | 38.491160 | 846.841774 |
| min | 9.000000 | 3.000000 | 68.000000 | 46.000000 | 1613.000000 |
| 25% | 17.500000 | 4.000000 | 104.250000 | 75.000000 | 2223.750000 |
| 50% | 23.000000 | 4.000000 | 148.500000 | 93.500000 | 2803.500000 |
| 75% | 29.000000 | 8.000000 | 262.000000 | 126.000000 | 3608.000000 |
| max | 46.600000 | 8.000000 | 455.000000 | 230.000000 | 5140.000000 |

| type | acceleration | model-year | origin |
|--------------|---------------------|-------------------|---------------|
| count | 398.000000 | 398.000000 | 398.000000 |
| mean | 15.568090 | 76.010050 | 1.572864 |
| std | 2.757689 | 3.697627 | 0.802055 |
| min | 8.000000 | 70.000000 | 1.000000 |
| 25% | 13.825000 | 73.000000 | 1.000000 |
| 50% | 15.500000 | 76.000000 | 1.000000 |
| 75% | 17.175000 | 79.000000 | 2.000000 |
| max | 24.800000 | 82.000000 | 3.000000 |

De todas las variables solo “horsepower” contiene datos faltantes seis en total. En los campos numéricos, el valor mínimo de cada uno es mayor que cero por lo que se intuye que no hay valores faltantes. Cada variable contiene su media, desviación estándar, etc.

Taxonomía de flores

Los primeros tres datos del conjunto

| sepal-length | sepal-width | petal-length | petal-width | class |
|--------------|-------------|--------------|-------------|-------------|
| 18.0 | 5.1 | 3.5 | 1.4 | Iris-setosa |
| 15.0 | 4.9 | 3.0 | 1.4 | Iris-setosa |
| 18.0 | 4.7 | 3.2 | 1.3 | Iris-setosa |

Informacion del data frame

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 150 entries, 0 to 149

Data columns (total 5 columns):

sepal-length 150 non-null float64

sepal-width 150 non-null float64

petal-length 150 non-null float64

petal-width 150 non-null float64

class 150 non-null object

dtypes: float64(4), object(1)

memory usage: 5.9+ KB

None

El conjunto de datos contiene 150 instancias, cuatro de las variables son numéricas y una es string.

| type | sepal-length | sepal-width | petal-length | petal-width |
|--------------|---------------------|--------------------|---------------------|--------------------|
| count | 150.000000 | 150.000000 | 150.000000 | 150.000000 |
| mean | 5.843333 | 3.054000 | 3.758667 | 1.198667 |
| std | 0.828066 | 0.433594 | 1.764420 | 0.763161 |
| min | 4.300000 | 2.000000 | 1.000000 | 0.100000 |
| 25% | 5.100000 | 2.800000 | 1.600000 | 0.300000 |
| 50% | 5.800000 | 3.000000 | 4.350000 | 1.300000 |
| 75% | 6.400000 | 3.300000 | 5.100000 | 1.800000 |
| max | 7.900000 | 4.400000 | 6.900000 | 2.500000 |

Como todas las variables contienen un mínimo mayor que cero se intuye que no hay valores faltantes, cada variable contiene su media, desviación estándar, etc.