

IPA 주관 인공지능센터 기본(fundamental) 과정

- GitHub link: [here](#)
- E-Mail: windkyle7@gmail.com

In [1]:

```
import pandas as pd
from sklearn.datasets import load_wine

wine = load_wine()
data = pd.DataFrame(wine.data, columns=wine.feature_names)
target = pd.DataFrame(wine.target, columns=['target'])
dataset = pd.concat([data, target], axis=1)
```

In [2]:

```
dataset.info()
dataset.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
alcohol                178 non-null float64
malic_acid             178 non-null float64
ash                   178 non-null float64
alcalinity_of_ash      178 non-null float64
magnesium              178 non-null float64
total_phenols          178 non-null float64
flavanoids             178 non-null float64
nonflavanoid_phenols  178 non-null float64
proanthocyanins        178 non-null float64
color_intensity        178 non-null float64
hue                   178 non-null float64
od280/od315_of_diluted_wines 178 non-null float64
proline                178 non-null float64
target                178 non-null int64
dtypes: float64(13), int64(1)
memory usage: 19.5 KB
```

Out[2]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	no
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	
mean	13.000618	2.336348	2.366517	19.494944	99.741573	2.295112	2.029270	
std	0.811827	1.117146	0.274344	3.339564	14.282484	0.625851	0.998859	
min	11.030000	0.740000	1.360000	10.600000	70.000000	0.980000	0.340000	
25%	12.362500	1.602500	2.210000	17.200000	88.000000	1.742500	1.205000	
50%	13.050000	1.865000	2.360000	19.500000	98.000000	2.355000	2.135000	
75%	13.677500	3.082500	2.557500	21.500000	107.000000	2.800000	2.875000	

max	14.830000	5.800000	3.230000	30.000000	162.000000	3.880000	5.080000	nc
	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	

In [3]:

```
import pandas_profiling as pp
profile = pp.ProfileReport(dataset)
profile
```

Out[3]:

Overview

Dataset info

Number of variables	14
Number of observations	178
Total Missing (%)	0.0%
Total size in memory	19.5 KiB
Average record size in memory	112.4 B

Variables types

Numeric	14
Categorical	0
Boolean	0
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

- `target` has 59 / 33.1% zeros Zeros

Variables

alcalinity_of_ash

Numeric

Distinct count	63
Unique (%)	35.4%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	19.495
Minimum	10.6
Maximum	30
Zeros (%)	0.0%



[Toggle details](#)

alcohol

Numeric

Distinct count	126
Unique (%)	70.8%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	13.001
Minimum	11.03
Maximum	14.83
Zeros (%)	0.0%



[Toggle details](#)

ash

Numeric

Distinct count	79
Unique (%)	44.4%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.3665
Minimum	1.36
Maximum	3.23
Zeros (%)	0.0%

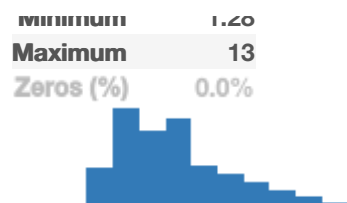


[Toggle details](#)

color_intensity

Numeric

Distinct count	132
Unique (%)	74.2%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	5.0581
Minimum	1.00



[Toggle details](#)

flavanoids

Numeric

Distinct count	132
Unique (%)	74.2%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.0293
Minimum	0.34
Maximum	5.08
Zeros (%)	0.0%



[Toggle details](#)

hue

Numeric

Distinct count	78
Unique (%)	43.8%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.95745
Minimum	0.48
Maximum	1.71
Zeros (%)	0.0%

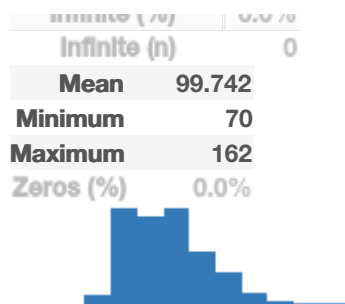


[Toggle details](#)

magnesium

Numeric

Distinct count	53
Unique (%)	29.8%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0



[Toggle details](#)

malic_acid

Numeric

Distinct count	133
Unique (%)	74.7%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.3363
Minimum	0.74
Maximum	5.8
Zeros (%)	0.0%



[Toggle details](#)

nonflavanoid_phenols

Numeric

Distinct count	39
Unique (%)	21.9%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.36185
Minimum	0.13
Maximum	0.66
Zeros (%)	0.0%

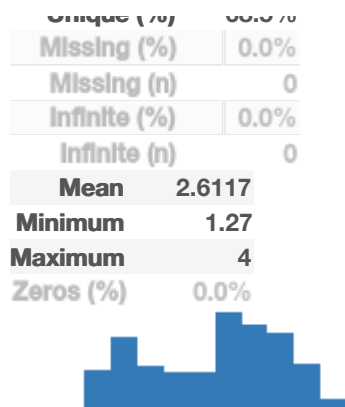


[Toggle details](#)

od280/od315_of_diluted_wines

Numeric

Distinct count	122
Unique (%)	68.5%



[Toggle details](#)

proanthocyanins

Numeric

Distinct count	101
Unique (%)	56.7%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	1.5909
Minimum	0.41
Maximum	3.58
Zeros (%)	0.0%



[Toggle details](#)

proline

Numeric

Distinct count	121
Unique (%)	68.0%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	746.89
Minimum	278
Maximum	1680
Zeros (%)	0.0%



[Toggle details](#)

target

Numeric

Distinct count	3
Unique (%)	1.7%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	0.9382
Minimum	0
Maximum	2
Zeros (%)	33.1%



[Toggle details](#)

total_phenols

Numeric

Distinct count	97
Unique (%)	54.5%
Missing (%)	0.0%
Missing (n)	0
Infinite (%)	0.0%
Infinite (n)	0
Mean	2.2951
Minimum	0.98
Maximum	3.88
Zeros (%)	0.0%



[Toggle details](#)

Correlations

Sample

	alcohol	malic_acid	ash	alkalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavonoids
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	
3	14.37	1.05	2.50	16.8	113.0	2.85	2.40	

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflav:
4	13.24	2.59	2.87	21.0	118.0	2.80	2.69	



In [4]:

```
with open('profile.html', 'w', encoding='utf8') as html:
    html.write(profile.to_html())
```

In []:

```
%%javascript
IPython.notebook.kernel.restart();
```

In [1]:

```
import pandas as pd
from sklearn.datasets import load_wine

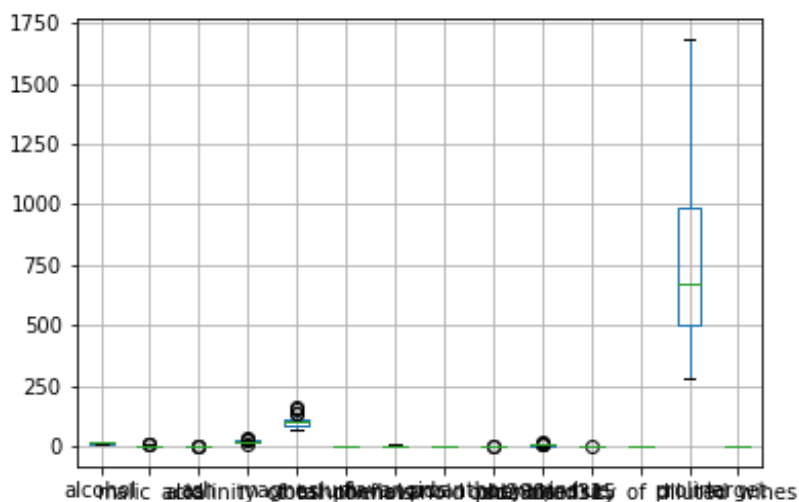
wine = load_wine()
data = pd.DataFrame(wine.data, columns=wine.feature_names)
target = pd.DataFrame(wine.target, columns=['target'])
dataset = pd.concat([data, target], axis=1)
```

In [2]:

```
%matplotlib inline
dataset.boxplot()
```

Out[2]:

<matplotlib.axes._subplots.AxesSubplot at 0x7fd6a47d4ba8>

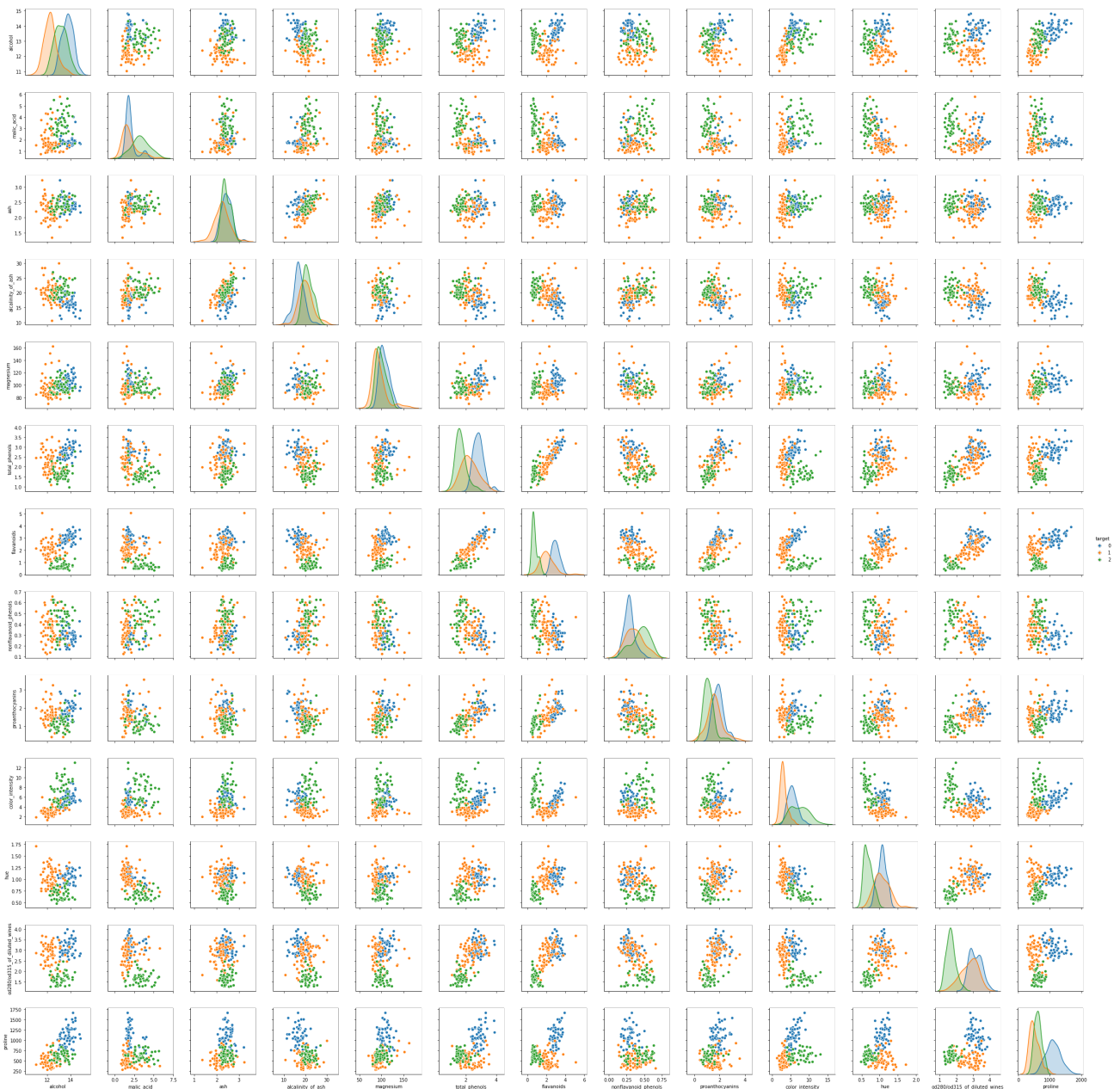



```
In [3]:
```

```
import seaborn as sns
sns.pairplot(dataset, vars=dataset.columns[:-1], hue='target')
```

```
Out[3]:
```

```
<seaborn.axisgrid.PairGrid at 0x7fd6a43cf048>
```



```
In [4]:
```

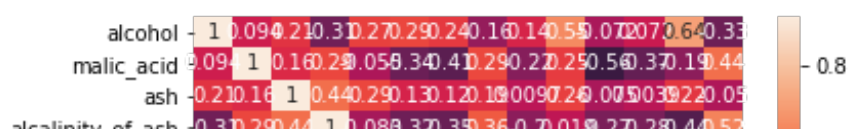
```
dataset.proline = dataset.proline.map(lambda x: (x-dataset.proline.mean())/
dataset.proline.std())
```

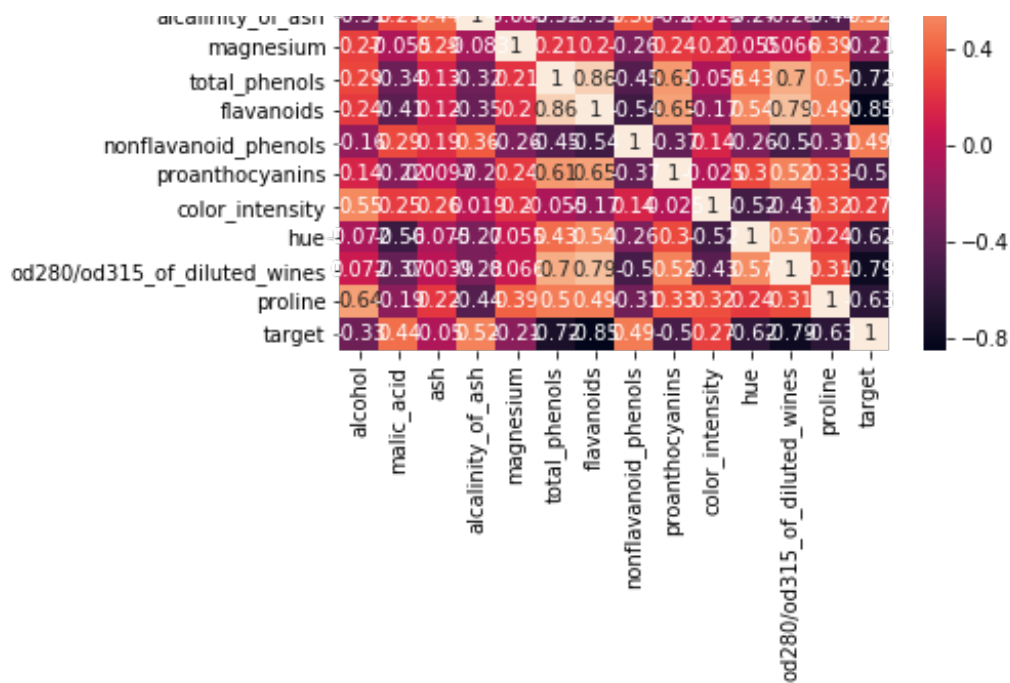
```
In [5]:
```

```
sns.heatmap(dataset.corr(), annot=True)
```

```
Out[5]:
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd699470c88>
```





In [6]:

```
from sklearn.decomposition import PCA

pca = PCA()
X = pd.DataFrame(pca.fit_transform(dataset.iloc[:, :-1]))
y = dataset.target
```

In [7]:

```
from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier

cross_val_score(KNeighborsClassifier(), X, y, cv=10)
```

Out[7]:

```
array([0.73684211, 0.88888889, 0.77777778, 0.94444444, 0.88888889,
       0.94444444, 0.72222222, 1.          , 0.76470588, 0.875       ])
```