indians

December 21, 2021

[ ]:

```
import numpy as np
import pandas as pd
ind_data = pd.read_csv("pima-indians-diabetes.csv", skiprows = 9, header = None)
ind_data
```

[ ]:

```
      0    1   2   3    4     5      6   7  8
0     6  148  72  35    0  33.6  0.627  50  1
1     1   85  66  29    0  26.6  0.351  31  0
2     8  183  64   0    0  23.3  0.672  32  1
3     1   89  66  23   94  28.1  0.167  21  0
4     0  137  40  35  168  43.1  2.288  33  1
..   ..  …   ..  ..  …    …     …    ..  ..
763  10  101  76  48  180  32.9  0.171  63  0
764   2  122  70  27    0  36.8  0.340  27  0
765   5  121  72  23  112  26.2  0.245  30  0
766   1  126  60   0    0  30.1  0.349  47  1
767   1   93  70  31    0  30.4  0.315  23  0

[768 rows x 9 columns]
```

1 /

- X = pima[,]
- y = pima[,]
- numpy index slicing

[ ]:

```
pima = ind_data.values
X = pima[:,:-1]
y = pima[:,-1]
pima.shape, X.shape, y.shape
```

[ ]:

```
((768, 9), (768, 8), (768,))
```

[ ]:

```
X = ind_data.iloc[:,:-1]
y = ind_data[8]
y = y.values
X.shape, y.shape
```

```
[ ]: ((768, 8), (768,))
```

```
[ ]: from sklearn.tree import DecisionTreeClassifier
     from sklearn.metrics import accuracy_score
     from sklearn.model_selection import train_test_split
```

```
[ ]: X_train, X_test, y_train, y_test = train_test_split(
         X, y, stratify = y,test_size = 0.2, random_state = 2021
     )
     X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
[ ]: ((614, 8), (154, 8), (614,), (154,))
```

2 GridSearchCV

```
[ ]: dtc = DecisionTreeClassifier(random_state=2021)
     dtc.fit(X_train, y_train)
     pred = dtc.predict(X_test)
```

```
[ ]: accuracy_score(y_test, pred), dtc.score(X_test, y_test)
```

```
[ ]: (0.7077922077922078, 0.7077922077922078)
```

```
[ ]: from sklearn.model_selection import GridSearchCV

     params = {
         'max_depth': [2,4,6,8],
         'min_samples_split': [2,4,6,8]
     }
     grid_dt = GridSearchCV(dtc, param_grid=params, scoring='accuracy', cv=3)
     grid_dt.fit(X_train, y_train)
```

```
[ ]: GridSearchCV(cv=3, estimator=DecisionTreeClassifier(random_state=2021),
                  param_grid={'max_depth': [2, 4, 6, 8],
                              'min_samples_split': [2, 4, 6, 8]},
                  scoring='accuracy')
```

```
[ ]: grid_dt.best_score_
```

```
[ ]: 0.7443248844253149
```

```
[ ]: grid_dt.best_params_
```

```
[ ]: {'max_depth': 2, 'min_samples_split': 2}
```

- Support Vector Machine

[ ]:

```
from sklearn.svm import SVC
svc = SVC(random_state = 2021)
svc.get_params()
params = {'C' : [3, 6, 10, 15, 20]}
grid_sv = GridSearchCV(svc, param_grid = params, scoring = "accuracy", cv = 5)
grid_sv.fit(X_train, y_train)
grid_sv.best_params_
```

```
[ ]: {'C': 3}
```

[ ]:

```
params = {'C' : [1, 2, 3]}
grid_sv = GridSearchCV(svc, param_grid = params, scoring = "accuracy", cv = 5)
grid_sv.fit(X_train, y_train)
grid_sv.best_params_
```

```
[ ]: {'C': 1}
```

[ ]:

```
params = {'C' : [0.01, 0.1, 0.5, 1]}
grid_sv = GridSearchCV(svc, param_grid=params, scoring='accuracy', cv=5)
grid_sv.fit(X_train, y_train)
grid_sv.best_params_
```

```
[ ]: {'C': 1}
```

[ ]:

```
best_svc = grid_sv.best_estimator_
best_svc.score(X_test, y_test)
```

```
[ ]: 0.7467532467532467
```