

# Рубежный контроль №1

Бахман Александр ИУ5-65Б

Импорт библиотек

```
In [1]: import pandas as pd
        from sklearn.preprocessing import LabelEncoder
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        from sklearn.datasets import load_iris
```

В качестве датасета возьмём данные Ирисов Фишера

```
In [13]: #Загрузка датасета и преобразование его в Dataframe
iris = load_iris()
data = pd.DataFrame(iris.data, columns=iris.feature_names)
data['target'] = pd.Series(iris.target) #создадим столбец целевой признака
data.head()
```

```
Out[13]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```
In [10]: data.shape
```

```
Out[10]: (150, 5)
```

В качестве целевого признака будем использовать признак "target"

```
In [11]: #Проверка на наличие пропусков
data.isnull().sum()
```

```
Out[11]: sepal length (cm)    0
sepal width (cm)           0
petal length (cm)          0
petal width (cm)           0
target                     0
dtype: int64
```

```
In [12]: #Проверка типов
data.dtypes
```

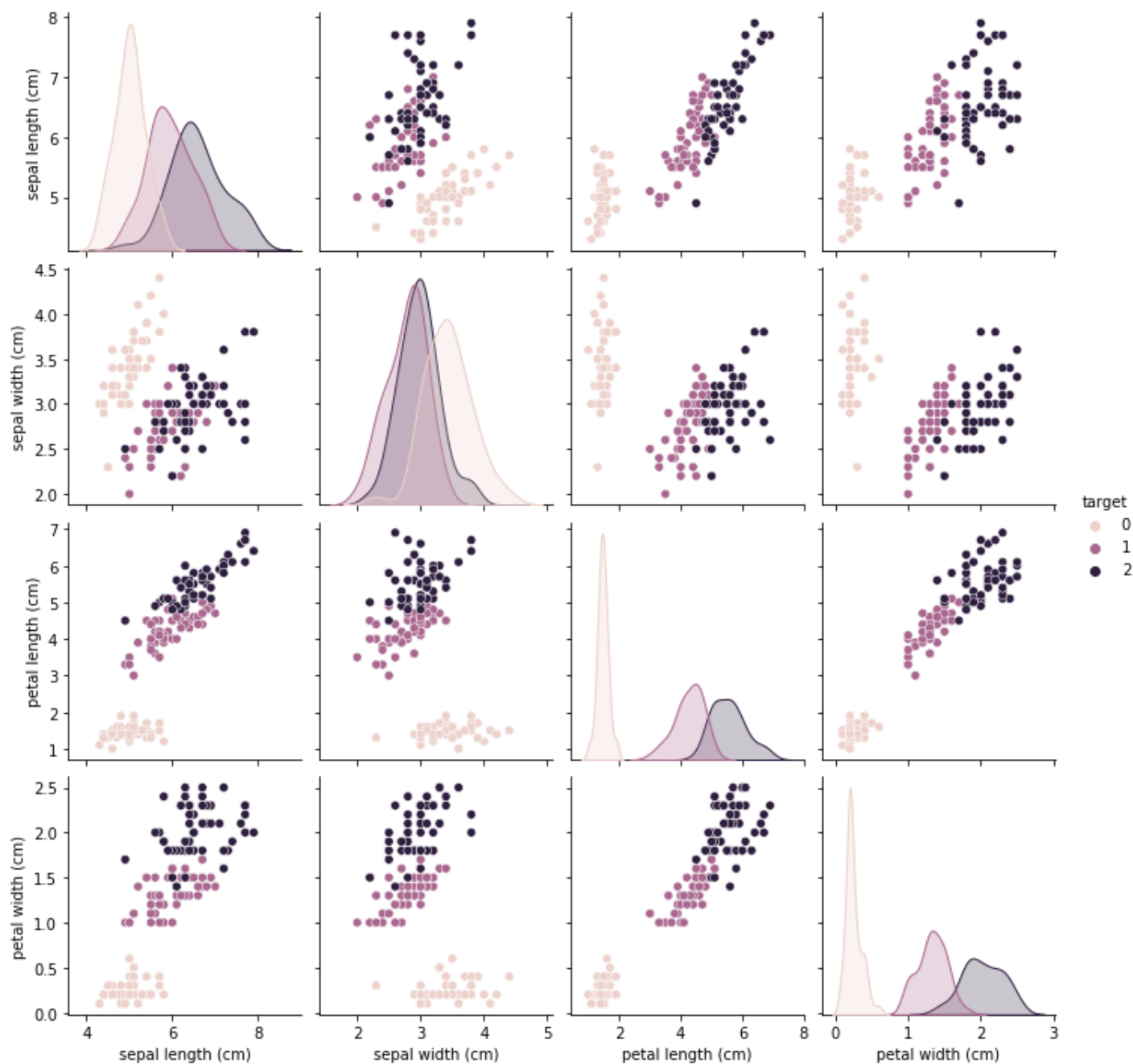
```
Out[12]: sepal length (cm)    float64
sepal width (cm)             float64
petal length (cm)            float64
petal width (cm)             float64
target                       int32
dtype: object
```

# Парные диаграммы

Построим парные диаграммы, выделив целевой признак

```
In [17]: sns.pairplot(data, hue="target")
```

```
Out[17]: <seaborn.axisgrid.PairGrid at 0x19a7695bfa0>
```



## Корреляционный анализ

Создадим корреляционную матрицу используя коэффициент Пирсона

```
In [18]: data.corr(method="pearson")
```

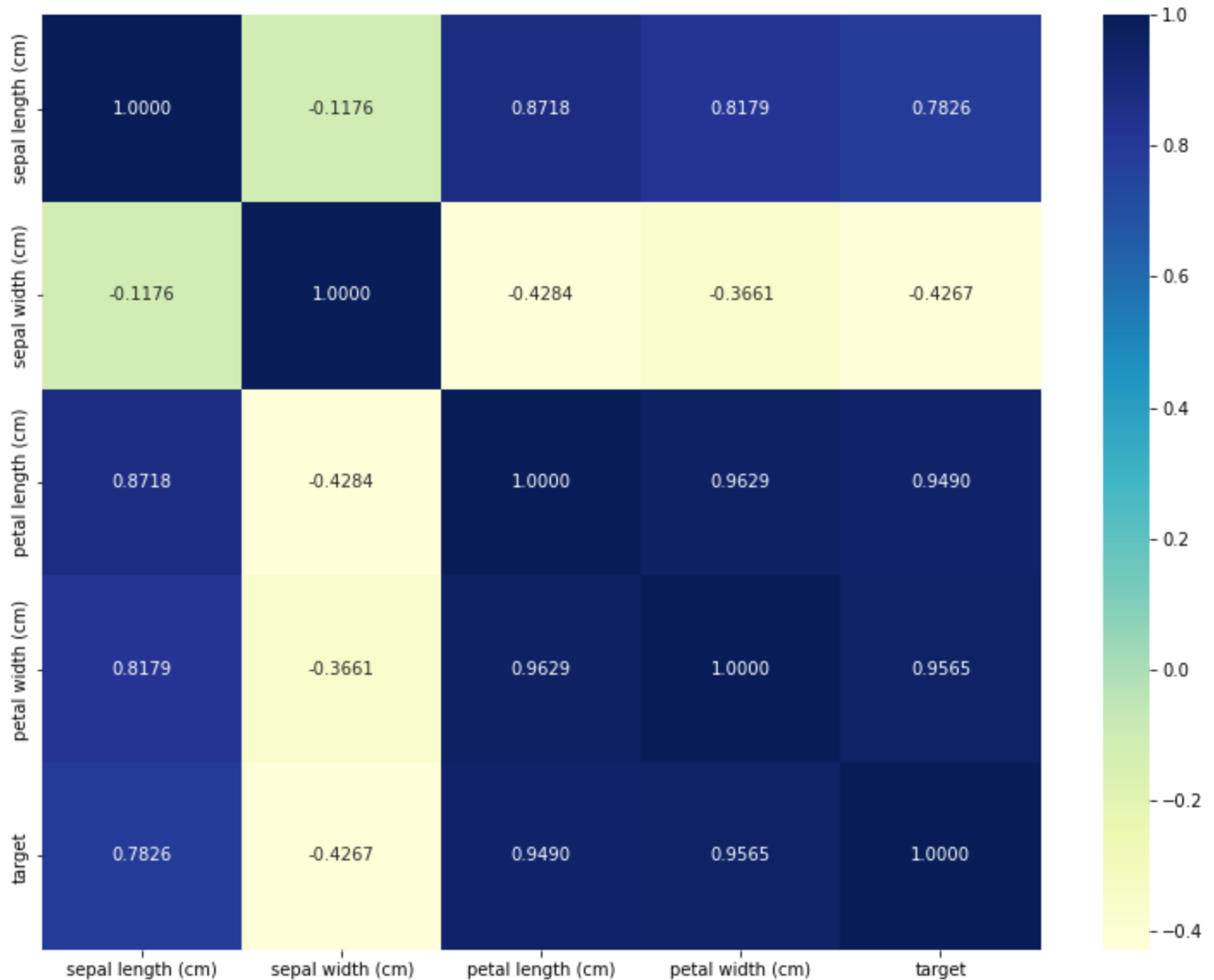
```
Out[18]:
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941	0.782561
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126	-0.426658

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
petal length (cm)	0.871754	-0.428440	1.000000	0.962865	0.949035
petal width (cm)	0.817941	-0.366126	0.962865	1.000000	0.956547
target	0.782561	-0.426658	0.949035	0.956547	1.000000

```
In [19]: fig, ax = plt.subplots(1,1, figsize=(13,10))
sns.heatmap(data.corr("pearson"), annot=True, fmt=".4f", cmap="YlGnBu")
```

```
Out[19]: <AxesSubplot:>
```



На основе корреляционной матрицы можно сделать следующие выводы:

Все признаки сильно коррелируют с целевым признаком Target, кроме признака sepal width. Признаки Petal lenght и Petal width сильно коррелируют между собой. При обучении модели эти признаки брать не стоит. Наилучшим выбором для обучения будет признак sepal width.

```
In [ ]:
```