

МИНОБРНАУКИ РОССИИ

Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

ОТЧЁТ
по практической работе №1
по дисциплине «Машинное обучение»

Студент гр. 6307

Гарифуллин В.Ф.

Преподаватель

Жангиров Т.Р.

Ход работы

Задание 1

Предположим X и Y две случайные переменные отражающие возраст и вес, соответственно. Рассмотрим случайную выборку из 20 наблюдений

X = (69, 74, 68, 70, 72, 67, 66, 70, 76, 68, 72, 79, 74, 67, 66, 71, 74, 75, 75, 76)

Y = (153, 175, 155, 135, 172, 150, 115, 137, 200, 130, 140, 265, 185, 112, 140, 150, 165, 185, 210, 220)

```
#1
X = np.array([69, 74, 68, 70, 72, 67, 66, 70, 76, 68, 72, 79, 74, 67, 66, 71, 74, 75, 75, 76])
Y = np.array([153, 175, 155, 135, 172, 150, 115, 137, 200, 130, 140, 265, 185, 112, 140, 150, 165, 185, 210, 220])
```

А. Найти среднее, медиану и моду величины X

```
#A
print(np.mean(X))      71.45
print(np.median(X))    71.5
print(stats.mode(X))   ModeResult(mode=array([74]), count=array([3]))
```

Среднее = 71.45

Медиана = 71.5

Мода = 74 (встречается 3 раза)

В. Найти дисперсию Y

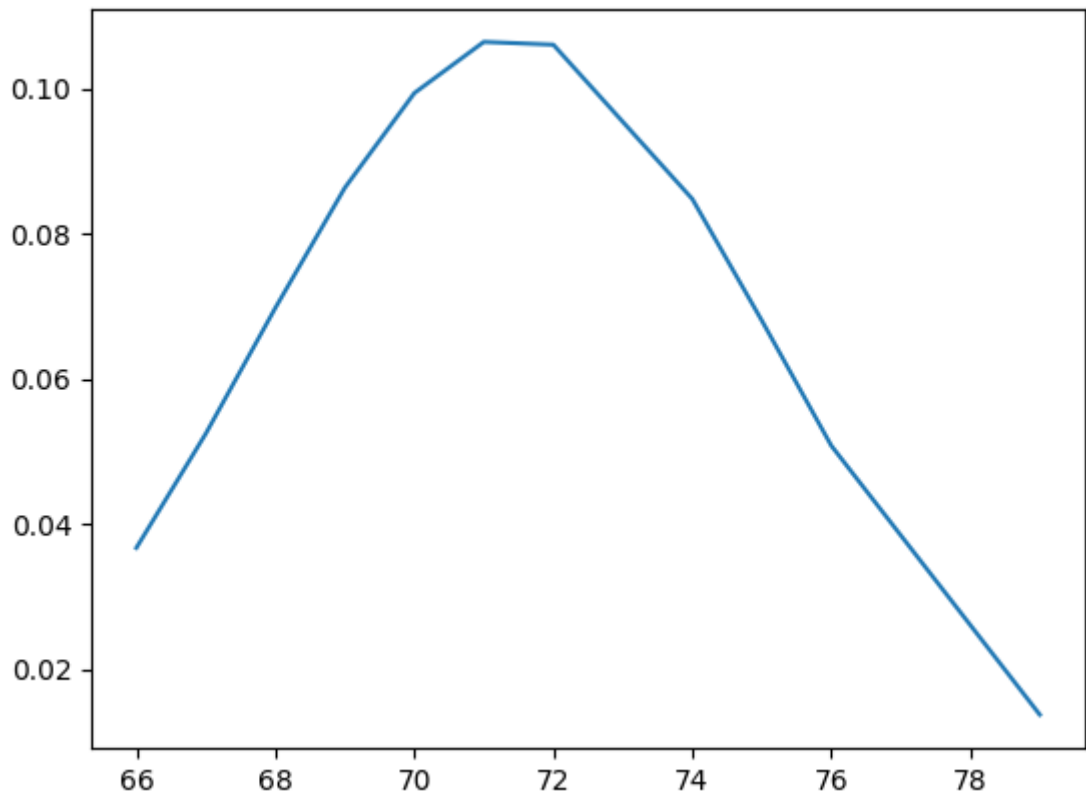
```
#B
print(np.var(Y))      1369.2099999999998
```

Дисперсия = 1369.21

С. Построить график нормального распределения для X

```
#C
XSorted = sorted(X)
plt.plot(XSorted, stats.norm.pdf(XSorted, np.mean(XSorted), np.std(XSorted)))
plt.show()
```

Сортируем X, рассчитываем МО и СКО, передаём в функцию, которая на их основе рассчитывает распределение вероятностей, строим по полученным данным график



D. Найти вероятность того, что возраст больше 80

```
#D
print(1 - stats.norm(np.mean(XSorted), np.std(XSorted)).cdf(80))
```

Рассчитываем кумулятивную функцию вероятности для 80, которая показывает какова вероятность, что возраст меньше 80. Вычитаем полученное число из единицы.

```
0.010791377919371459
```

Вероятность того, что возраст больше 80 = 0.0108

Е. Найти двумерное мат. ожидания и ковариационную матрицу для этих двух величин

Мат. ожидание:

```
#E
print(np.mean([X, Y], axis=1))
```

```
[ 71.45 164.7 ]
```

Ковариационная матрица:

```
print(np.cov(X, Y))
```

```
[[ 14.57631579 128.87894737]
 [128.87894737 1441.27368421]]
```

F. Определять корреляцию между X и Y

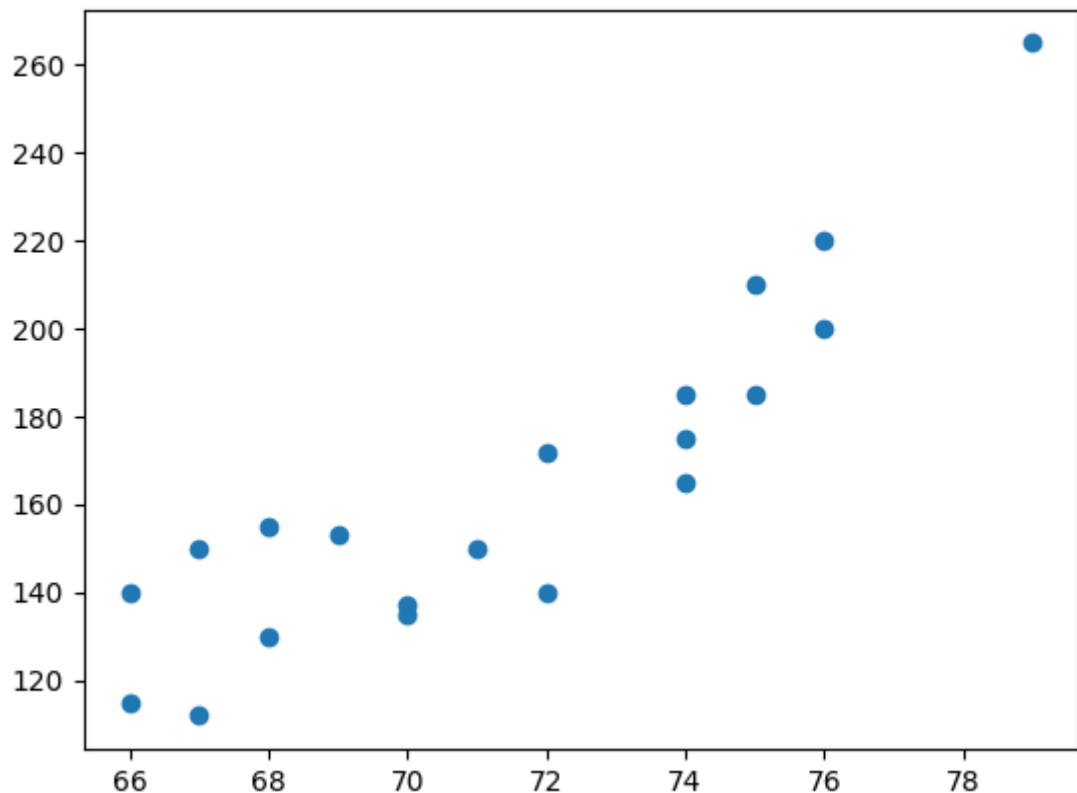
```
#F
print(np.corrcoef(X, Y)[1, 0])
```

```
0.8891701351748048
```

Корреляция = 0.889

G. Построить диаграмму рассеяния, отображающая зависимость между возрастом и весом

```
#G
plt.scatter(X, Y)
plt.show()
```



Задание 2

Для следующего набора данных

	X_1	X_2	X_3
a	17	17	12
b	11	9	13
c	11	8	19

```
#2  
arr = np.array([[17, 11, 11], [17, 9, 8], [12, 13, 19]])
```

Рассчитайте ковариационную матрицу и обобщенную дисперсию

Ковариационная матрица:

```
[[ 12.          17.         -8.          ]  
 [ 17.          24.33333333 -12.83333333]  
 [ -8.          -12.83333333  14.33333333]]
```

Обобщённая дисперсия рассчитывается, как детерминант ковариационной матрицы:

```
print(np.linalg.det(np.cov(arr)))
```

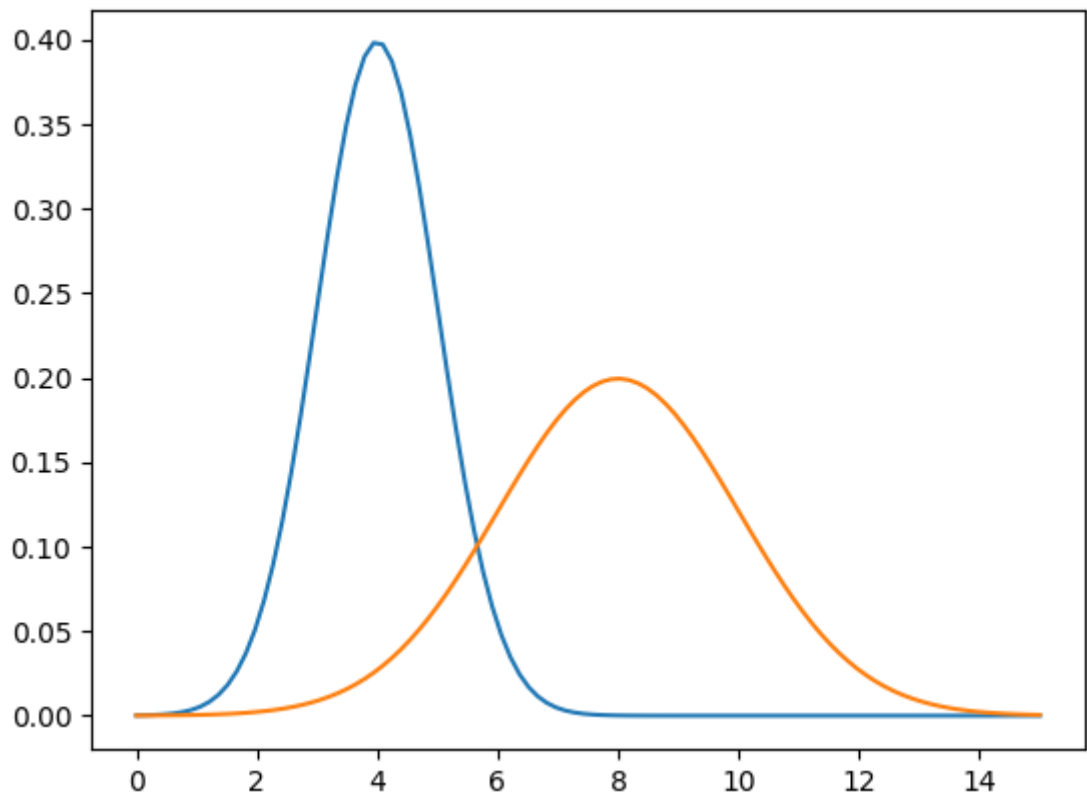
```
9.577387902356475e-14
```

Обобщённая дисперсия = $9.577e-14$

Задание 3

Даны два одномерных нормальных распределения N_a и N_b с мат. ожиданиями 4, 8 и СКО 1, 2 соответственно.

```
#3  
pltX = np.linspace(0, 15, 100)  
plt.plot(pltX, stats.norm.pdf(pltX, 4, 1))  
plt.plot(pltX, stats.norm.pdf(pltX, 8, 2))  
plt.show()
```



А. Для каждого из значения $\{5,6,7\}$ определите какое из распределений сгенерировало значение с большей вероятностью.

5 – Na, 6 – Nb, 7 – Nb

В. Найди значение, которой могло быть сгенерировано обеими распределениями с равной вероятностью

5.66 с вероятностью 0.1