1. Даны значения величины заработной платы заемщиков банка (zp) и значения их поведенческого кредитного скоринга (ks): zp = [35, 45, 190, 200, 40, 70, 54, 150, 120, 110], ks = [401, 574, 874, 919, 459, 739, 653, 902, 746, 832]. Используя математические операции, посчитать коэффициенты линейной регрессии, приняв за X заработную плату (то есть, zp - признак), а за y - значения скорингового балла (то есть, ks - целевая переменная). Произвести расчет как с использованием intercept,

n = 10  
  
x = np.array([35, 45, 190, 200, 40, 70, 54, 150, 120, 110], dtype=np.float64)  
y = np.array([401, 574, 874, 919, 459, 739, 653, 902, 746, 832], dtype=np.float64)  
  
b = (np.mean(x \* y) - np.mean(x) \* np.mean(y)) / (np.mean(x\*\*2) - np.mean(x) \*\* 2)

b = 2.620538882402765

a = np.mean(y) - b \* np.mean(x)

a = 444.1773573243596

так и без.

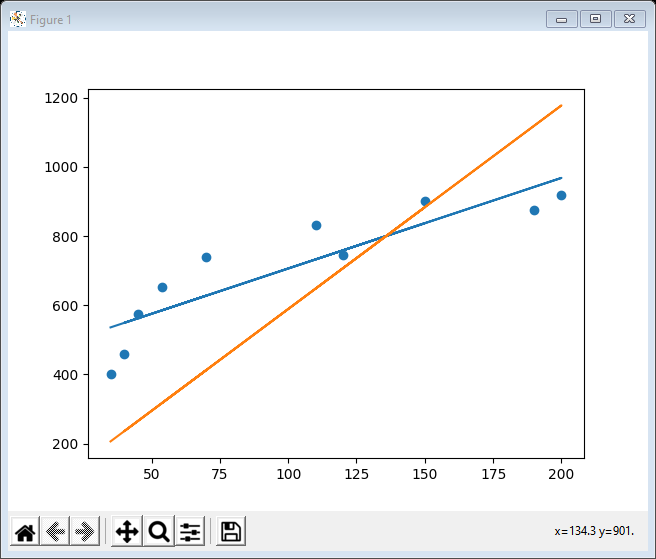
Sum (b\*x-y)^2 -> min сумма квадратов отклонений стремится к минимуму, производная = 0

Sum (2\*(b\*x-y)\*x) = sum (2b\*x^2 –2x\*y) = b\* sum(2\* x^2) –sum (2\*x\*y) = 0

B= sum (x\*y)/ \* sum( x^2)

b1 = np.sum(x\*y)/np.sum(x\*x)

без intercept b = 5.889820420132689



1. Посчитать коэффициент линейной регрессии при заработной плате (zp), используя градиентный спуск (без intercept).

iteration 500, B1 = 5.889815458885515, mse = 56516.858416058494

iteration 1000, B1 = 5.889820420128437, mse = 56516.8584157194

iteration 1500, B1 = 5.889820420132673, mse = 56516.85841571943

iteration 2000, B1 = 5.889820420132673, mse = 56516.85841571943

1. В каких случаях для вычисления доверительных интервалов и проверки статистических гипотез используется таблица значений функции Лапласа, а в каких - таблица критических точек распределения Стьюдента?

Таблица значений функции Лапласа используется при неизвестном математическом ожидании и известной дисперсии нормально распределенной случайной величины

Таблица критических точек распределения Стьюдента используется при неизвестном математическом ожидании и дисперсии нормально распределенной случайной величины