**Report Introduction to Data Science – Assignment 5**

**Linear regression**

**Exercise 1**

The weights for first-feature training set are: [ 5.20 0.05]

The weights for 12-features training set are: [5.16e+01 1.95e-02 -1.06e+00 2.58e-02 5.02e-02 -2.75e+00 5.65e-03 -3.80e-03 -4.72e+01 -4.26e-01 8.50e-01 2.37e-01]

**Exercise 2**

The rmse for one feature test set is: 0.78

The rmse for all features test set is: 0.64

**Random forest**

Using this normalization affects different classification methods differently. How does it affect a random forest classifier?

Short discussion on how normalization of the features to zero mean and unit variance affects a random forest

**Exercise 3**

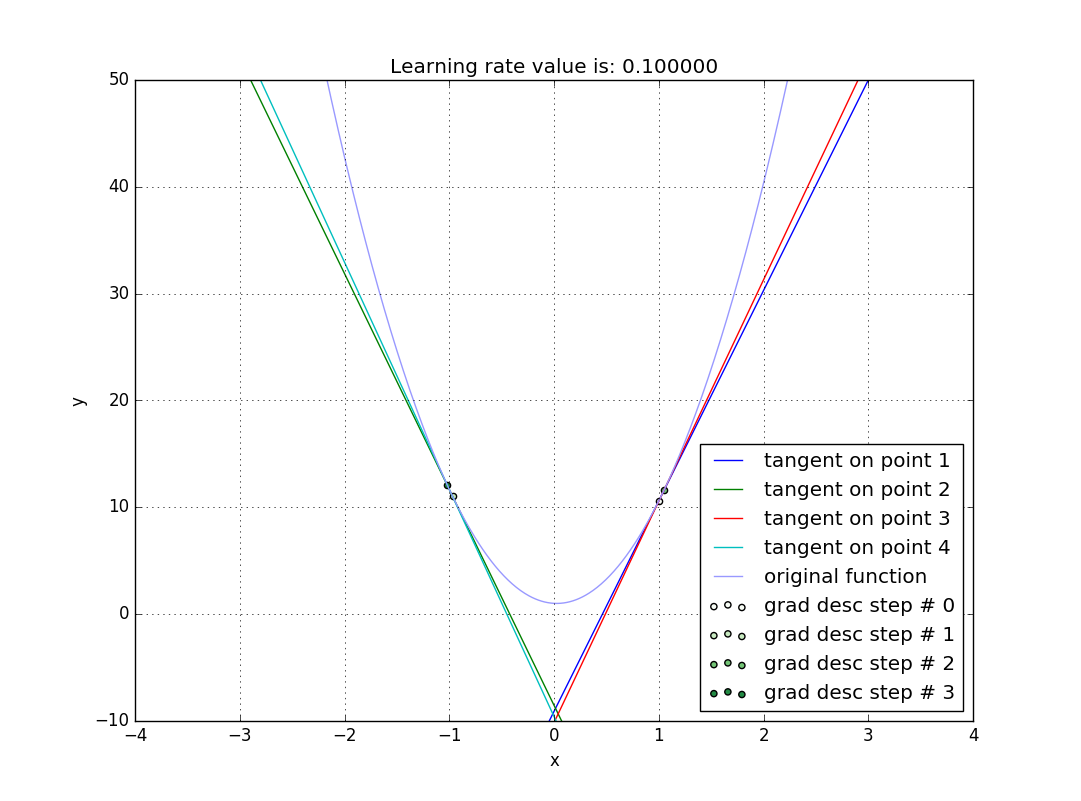
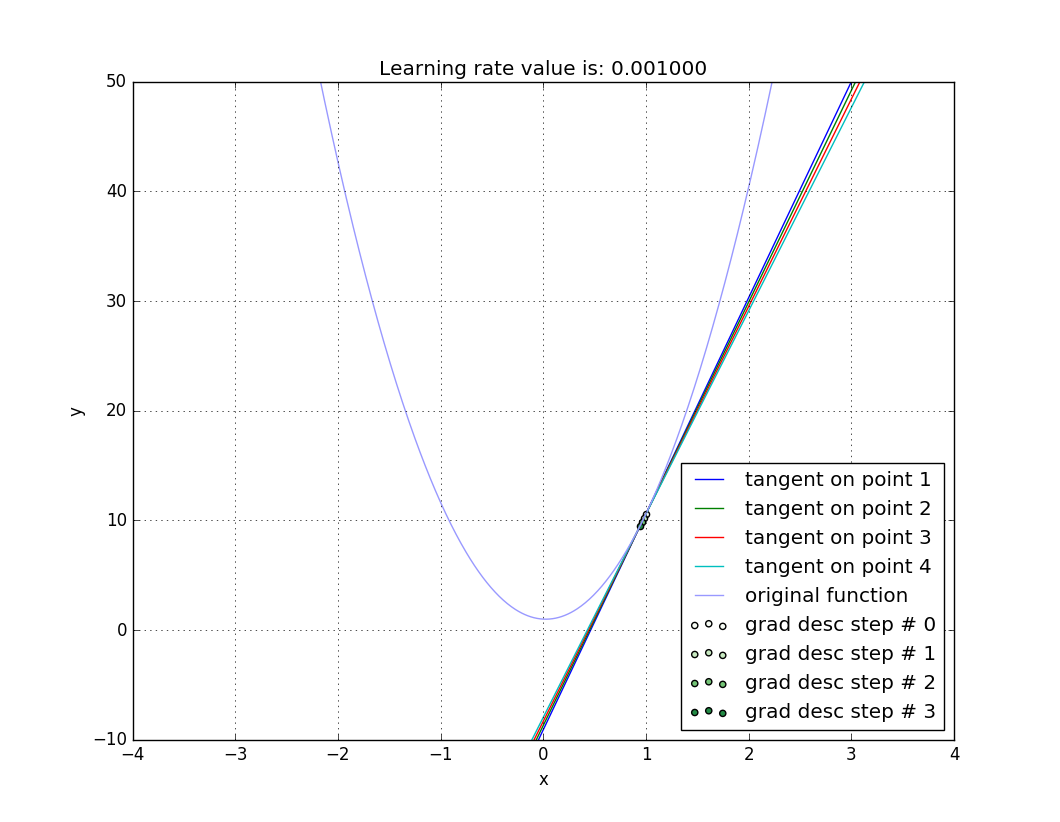
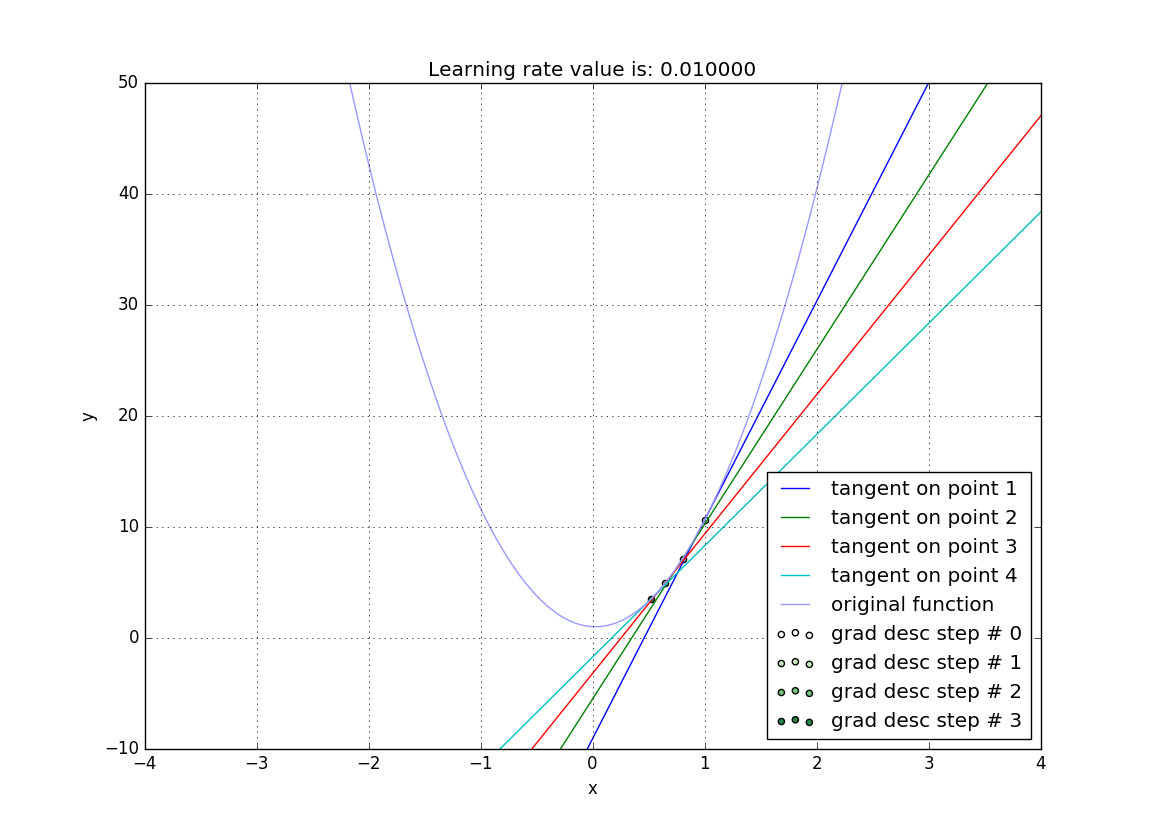
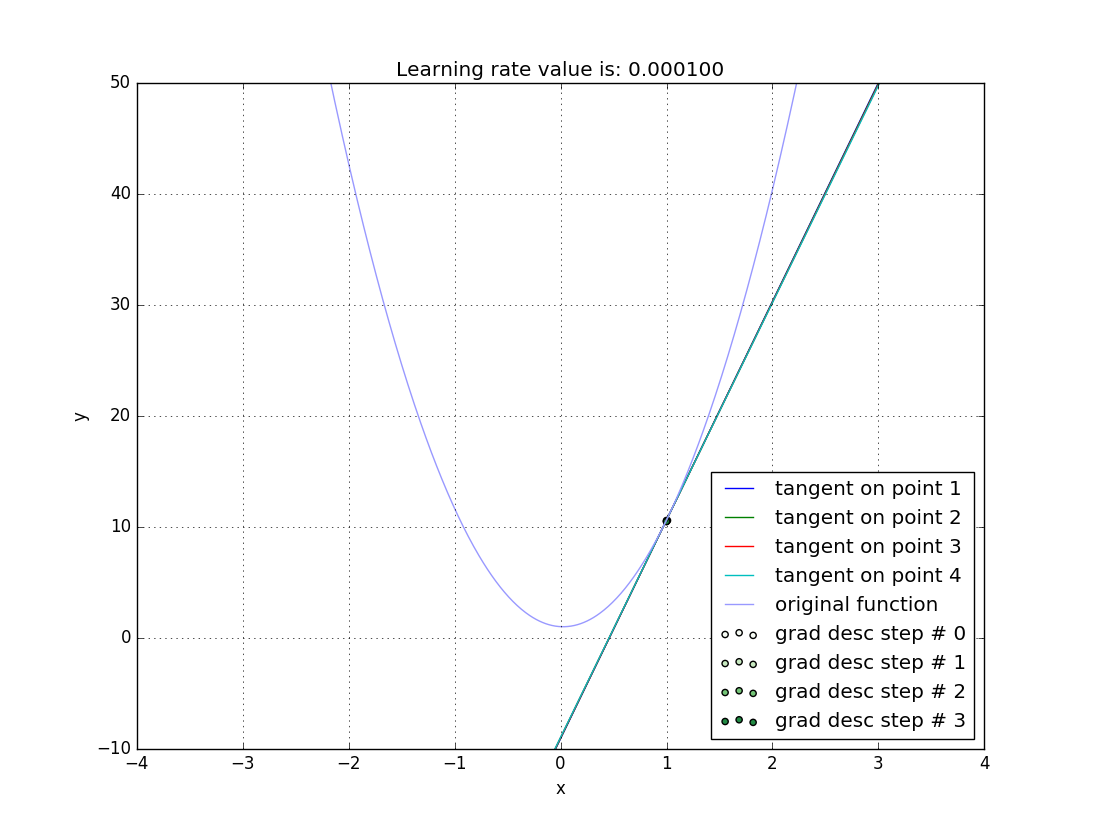
Does it beat the nearest neighbour classifier? Yes

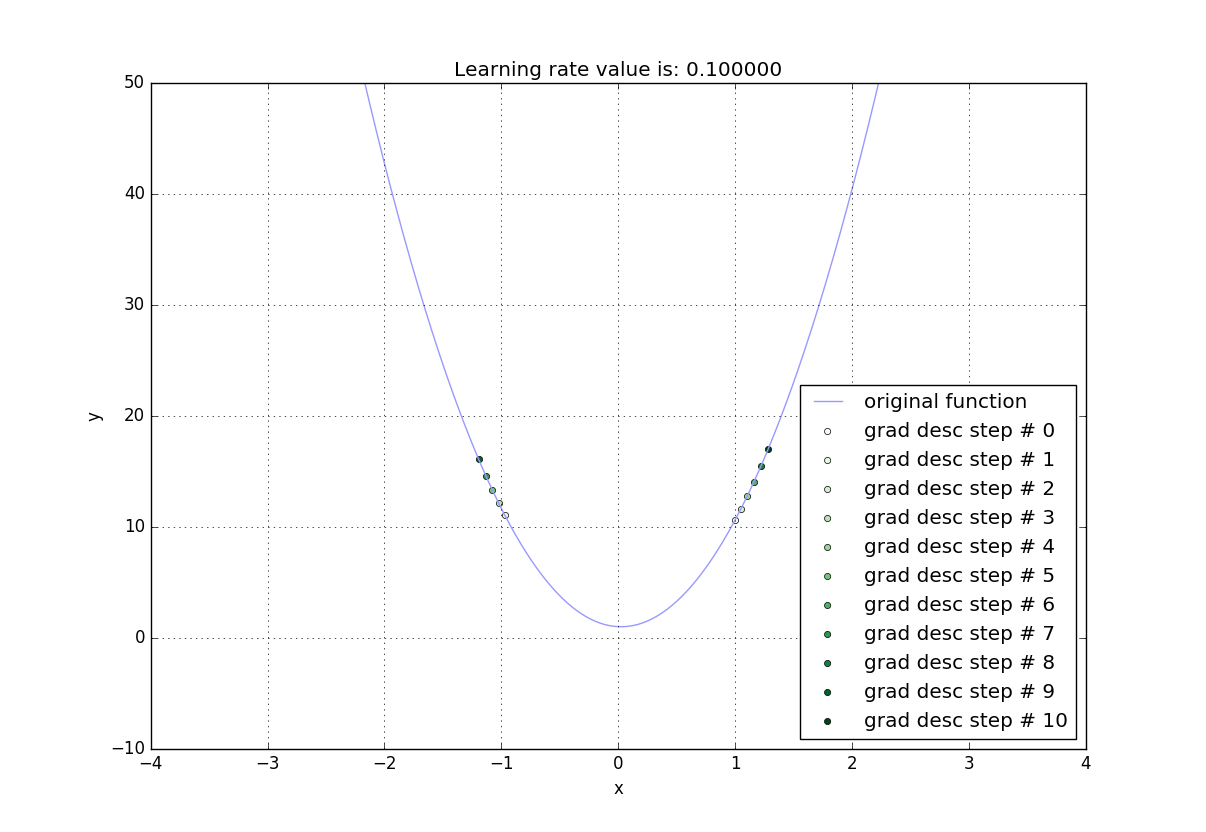
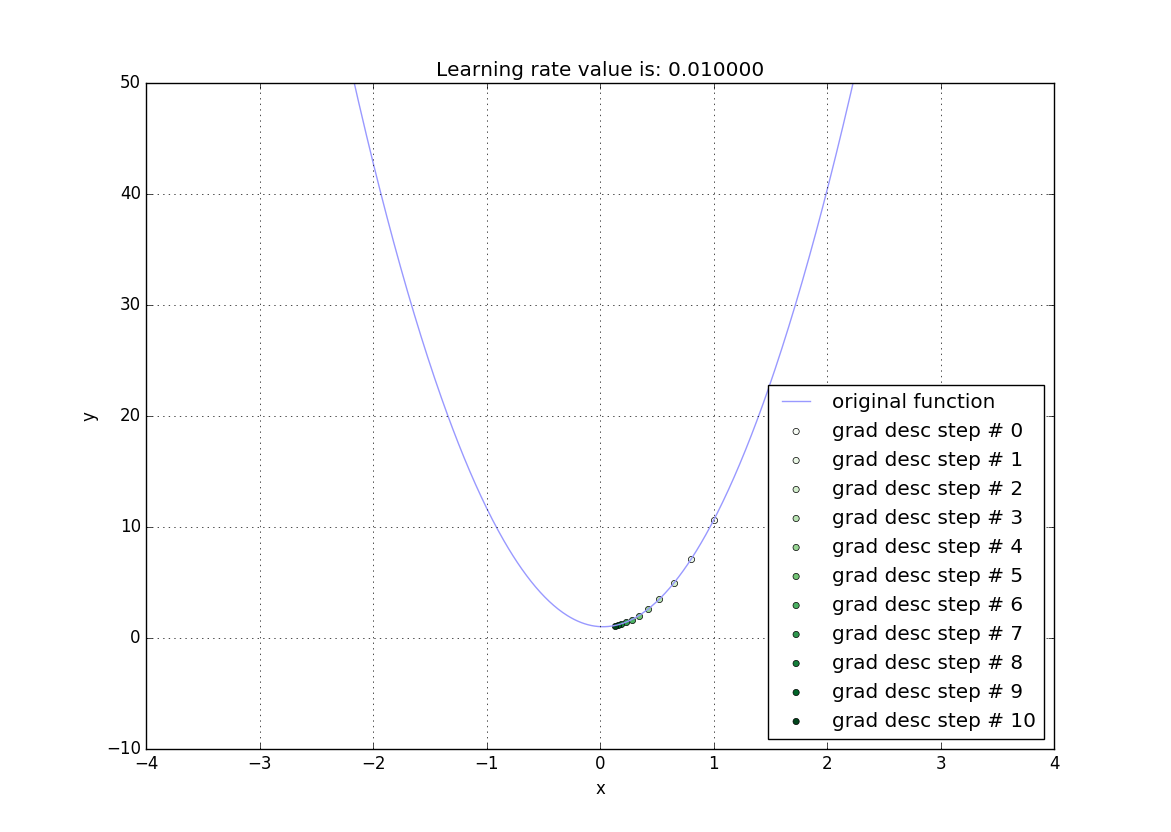
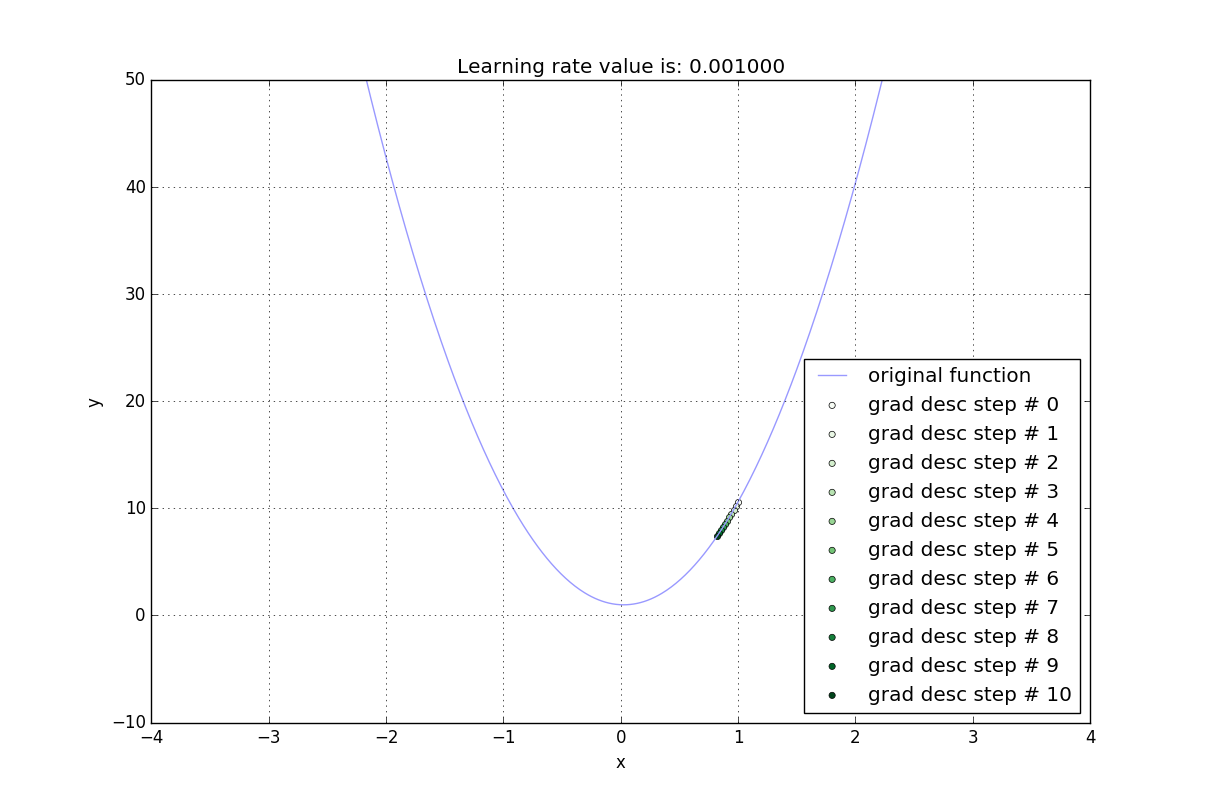
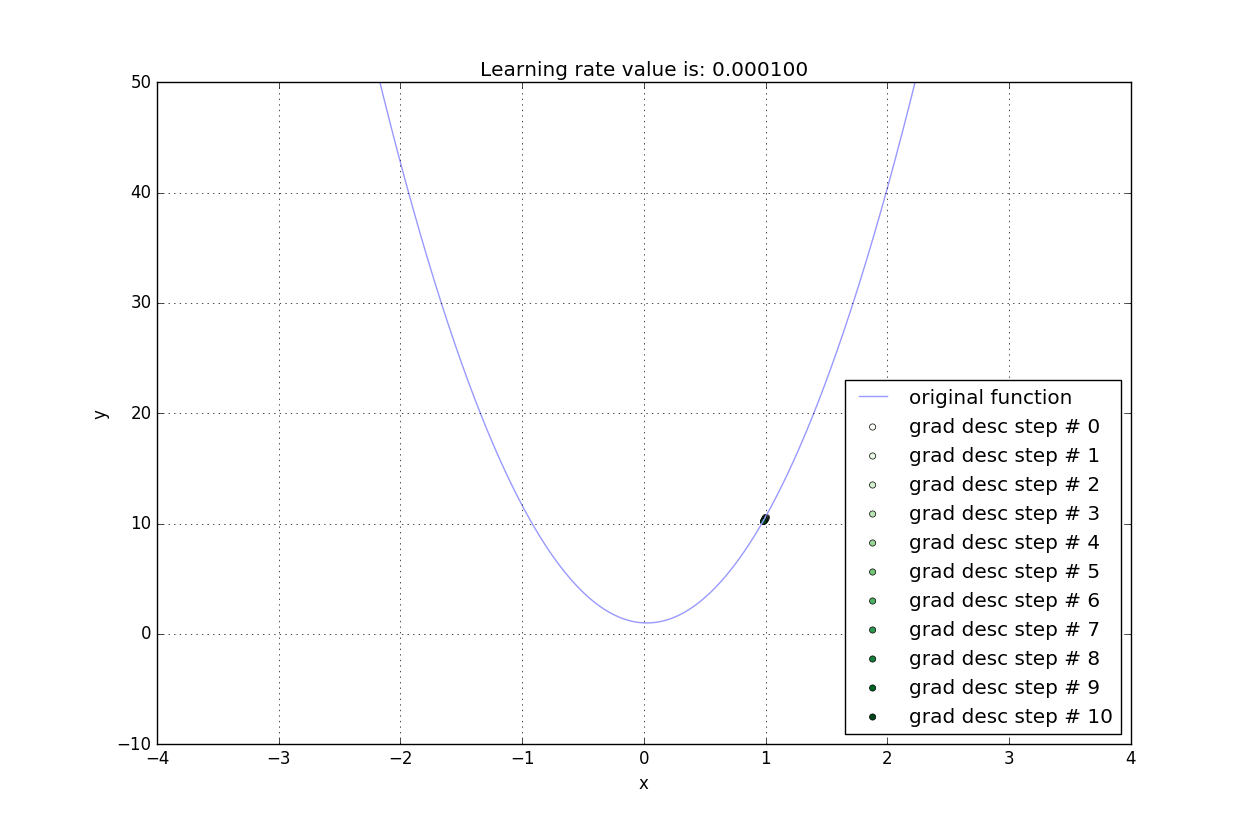
**Exercise 4**

The accuracy score for the test set is: 0.968

**Exercise 5**

**Gradient descent**

**b)**

**c)**

**d)**

**For learning rate 0.1:**

**File "<string>", line 1, in <lambda>**

**OverflowError: math range error**

**with learning rate: 0.01**

the function value in the last iteration is: 0.02469323271

the final iteration is iteration number: 95

**with learning rate: 0.001**

the function value in the last iteration is: 0.0246932372

the final iteration is iteration number: 934

**with learning rate: 0.0001**

the function value in the last iteration is: 0.0246932816

the final iteration is iteration number: 8290

**Exercise 6**

**Logistic Regression**