Databases Assignment 4

Part I Linear regression:

Added features: We have added two features, the “University” a number between 1 and 3, 1 for Ariel, 2 for Ivrit, 3 for Technion. And the “Nationality” a number between 1 and 3, 1 for Russia, 2 for France, 3 for Israel.
The features have been added arbitrary, without worried about the real. Which mean that adding this information doesn’t mean an improvement of the prediction of the salary.
As a result we conclude : adding features do not improve the error cause the features we add are not relevant, and not based on the real life.
We here show that wrong information are detected. See on the ScreenShots folder the screenshot which show the output of the linear regression and the two error.

Part II logistic regression:

added features: We have added two features, “Management” a binary indicator which is 1 when salary is >=30k and “Junior” when salary<=20k
We tried to make the features informative and helpful but the names\meanings are of course arbitrary (tough features added artificially are considered noise in our opinion more then features).
We have had mixed results regarding the helpfulness of the added features, probably because the data set was too small.

With normalization and without added features:

```
/usr/bin/python3.5 /home/ehud/PycharmProjects/clasifier/.idea/linear.py
100k iterations , LR=0.005 , weight vector was initialized with 0, 70% train data ,30% test data
TP:  8  TN:  3  FN:  3  FP:  3
accuracy:  0.7333333333333333  precision:  0.7272727272727273  recall:  0.7272727272727273  F1:  0.7272727272727273

Process finished with exit code 0
```

```
/usr/bin/python3.5 /home/ehud/PycharmProjects/clasifier/.idea/linear.py
100k iterations , LR=0.005 , weight vector was initialized with 0, 70% train data ,30% test data
TP:  9  TN:  3  FN:  2  FP:  2
accuracy:  0.8  precision:  0.8181818181818182  recall:  0.8181818181818182  F1:  0.8181818181818182

Process finished with exit code 0
```

with added features and normalization:

```
100k iterations , LR=0.005 , weight vector was initialized with 0, 70% train data ,30% test data
TP:  10  TN:  1  FN:  3  FP:  3
accuracy:  0.7333333333333333  precision:  0.7692307692307693  recall:  0.7692307692307693  F1:  0.7692307692307693
```

```
/usr/bin/python3.5 /home/ehud/PycharmProjects/clasifier/.idea/linear.py
100k iterations , LR=0.005 , weight vector was initialized with 0, 70% train data ,30% test data
TP:  7  TN:  2  FN:  4  FP:  4
accuracy:  0.6  precision:  0.6363636363636364  recall:  0.6363636363636364  F1:  0.6363636363636364

Process finished with exit code 0
```