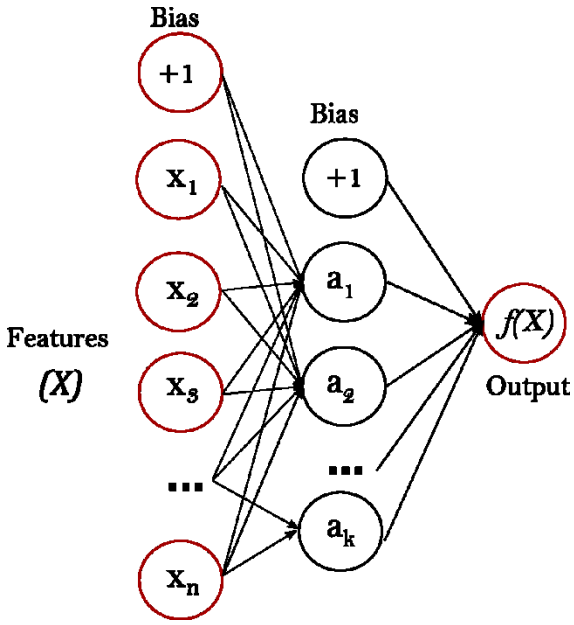
Introduction to Data Mining

Neural Networks

Question 1

Review the online document for neural networks in scikitlearn: http://scikit-learn.org/stable/modules/neural_networks_supervised.html.



Question 2 – Load Data

We will implement a neural network to predict customer churn. A dataset is provided in the file "bank_customer_churn.csv" – load it into your worksheet.

Features:

- Customer ID
- Surname
- CreditScore
- Geography
- Gender
- Age
- Tenure (years as a customer)
- Balance
- NumOfProducts
- HasCrCard
- IsActiveMember
- EstimatedSalary

Target Variable:

- Exited (i.e. left the bank / did churn)

There are 10,000 instances.

Question 3 – Feature Engineering

In this question we will prepare the features for using in a NN.

- Process the features to make them numeric (change categorical variables into features)
- Scale the variables to be in the same range (e.g. -1 to 1)
    - You can use sklearn.preprocessing and review StandardScalar
    - Also see the section containing tips on practical use and scaling in the NN docs

Question 4 – Create the training and test sets

Use train_test_spit

Question 5 – Implement the NN

Use a multi-layer perceptron NN (MLP) with Stochastic Gradient Descent (it differs from the basic gradient descent algorithm described in lectures by sometimes randomly jumping instead of selecting the steepest descent to attempt to avoid local minima).

Experiment with different activation functions and hidden layer configurations. Try to build a loop to test different parameters.

Question 6

Result reporting. Describe the performance of your model including accuracy, AUC, etc. Compare with a decision tree or logistic regression.