COMP8270
Programming for Artificial Intelligence

Class 12

In this practical class we will practice building a classifier with Keras. We will build a model to predict loan defaults on mortgages.

Create a new jupyter notebook and call it Santa.

We will use real data. You will need to load:

```
loan = pd.read_csv ("http://raptor.kent.ac.uk/~ds756/loan.csv")
```

There are simply too many predictors (features/columns) for us to finish fitting our model in the time allotted, so our first task is to drop some columns. Just cut & paste the following:

```
subLoan = loan.drop (columns=['ID', 'Funded Amount Investor',
    "Batch Enrolled",
    'Grade', 'Sub Grade',
    'Employment Duration', 'Home Ownership', 'Verification Status',
    'Payment Plan', 'Loan Title',
    'Delinquency - two years', 'Inquires - six months', 'Open Account',
    'Public Record', 'Revolving Balance',
    'Total Accounts', 'Initial List Status', 'Total Received Interest',
    'Total Received Late Fee', 'Recoveries', 'Collection Recovery Fee',
    'Collection 12 months Medical', 'Application Type', 'Last week Pay',
    'Accounts Delinquent', 'Total Collection Amount',
    'Total Current Balance'])
```

Steps:

1. Import Keras, and any others packages that you will need.
2. We want to use loan default as the category so create a new column in the Pandas, “Default”. Take the “Loan Status” feature and map it to the strings “Default” or “Serviced” using a lambda.

3. Drop the “Loan Status” column.
4. We now have a Pandas with 7 numeric predictors and 1 categorical feature.
5. Turn the Pandas into an array.
6. Split off the numeric predictors
7. Transform the categorical feature in to a hot-encoded vector.
8. Split your data in to validation and test subsets.
9. Create a ANN:
   Add a normalization layer
   2 layers with 8 neurons each
   A softmax layer
   compile it for classification

10. Fit the model. Try 1000 epochs and a batch size 500.
11. Try your validation data on your trained model. How did you?

If you have spare time, try plotting loss vs. batch size. You will have to fit your model for each point on your graph.

Have a great holiday!