Machine Learning and Finance – Final exam

A. Credit Risk Prediction: [70 marks]

We wish to create a model to predict the quality of a loan. Two categories are possible: good (G) refers to a low-risk level and bad (B) refer to a high-risk level.

In order to come with an assessment we may rely on the following features:

- Age, a numerical value between 18 and 100
- Geography, a string, either Germany, France, the UK, China, or Japan
- Sex, a string, male or female
- Purpose, a string, P1, P2, P3, P4
- Credit, a numerical value in $
- Duration, a numerical value in months
- Education, a string, unskilled, skilled

Convention: To binarize the targets, the category "good" will be considered as the positive target (1), and the category "bad" will be associated to the target (0).

A.1 Preprocessing

Before feeding the data into the Machine Learning model, we should preprocess it:

(1) Determine which of these features are categorical and describe the one hot encoding preprocessing method. [6 marks]

(2) What transformation should be applied to the numerical values to normalize them? [6 marks]

A.2 Logistic Regression

We wish to parameterize a logistic regression model on 1000 samples. We therefore split the dataset into 75% allocated to the train set and 25% to the test set.

After training the logistic regression model, we end up with the vector of optimal weights $\boldsymbol{w}^*$. The prediction of the target based on a new feature vector $x \in R^D$ can be expressed as follows:

$$P(Y = 1 \,|x) = \sigma(w^{*T} x)$$

Where sigma stands for the sigmoid function. We want to determine the optimal threshold value T* to turn the real valued probabilistic predictions into 0/1 decisions. We obtain the following results for three thresholds T1, T2, T3:

- T1: TPR=0.6; FPR=0.2
- T2: TPR=0.8; FPR=0.2
- T3: TPR=0.8; FPR=0.6

Where TPR stands for True Positive Rate and FPR stands for False Positive Rate.

(3) According to the previous results, which threshold should we use? Detail your answer [6 marks]

We now estimate three different logistic models: a linear logistic one, a quadratic one and a kernel logit. The first one delivers a Gini coefficient of 65%, the second 71% and the last one of 73%.

(4) Which model looks preferable? ***[6 marks]***

At the threshold T2 level, the lowest type 1 errors (Type 1 error (false positive) = the person is in default (0) but is predicted as a non defaulter (1)) are found for the quadratic logit.

(5) Given the retained threshold, which model should be used? [6 marks]

A.3 Decision Tree Algorithm

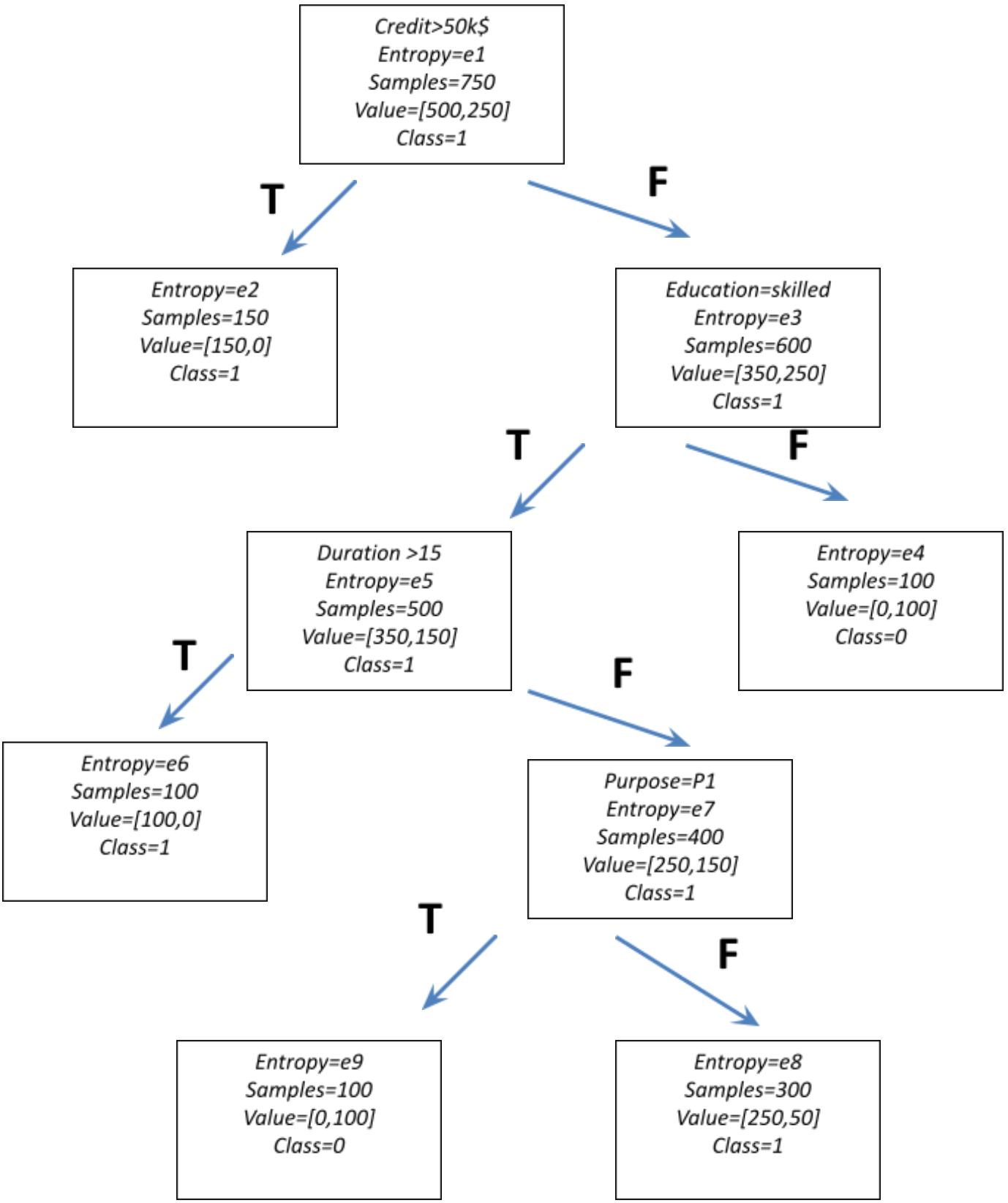
We wish to compare the logistic regression model with the Decision Tree Algorithm on the dataset. After training the Decision Tree model, we end up with the decision graph on the next page.

(6) What will a null entropy value correspond to on such a graph? [6 marks]

(7) In the decision graph, which of the entropies e_i, with i from 1 to 9, will be equal to zero? [6 marks]

(8) From the decision graph, deduce the confusion matrix. Justify your answer. [6 marks]

	1	0
1	TP	FN
0	FP	TN



(9) Explain why the precision is a good evaluation metric. [6 marks]

(10) Compare the precisions of the Decision Tree Algorithm and of the logistic regression associated to the T2 threshold (with TPR = 0.8 and FPR = 0.2) on the training data. [10 marks]

We have the following new sample with the following features:

Age: 24
Geography: UK
Sex: Female
Credit: 35k$
Education: skilled
Duration: 11

(11) What should be the value of "Purpose" to make the DT algorithm predict a 'Good' target? Justify your answer [6 marks]

B. <u>A simple Markov Model:</u> [15 marks]

Let's consider the following three sentences: `

S1: Deep Learning is part of Machine Learning.
S2: Machine Learning is part of Artificial Intelligence.
S3: Artificial Intelligence is fun.

In this example, our vocabulary is:

Vocabulary={"Deep", "Learning", "is", "part", "of", "Machine", "Artificial", "Intelligence", "fun"}

(12) What method would you use to learn the parameters of a Markov Model on this small dataset [3 marks]

We associate each word to an integer state as follows:

"Deep": 0, "Learning": 1, "is": 2, "part": 3, "of": 4, "Machine": 5, "Artificial": 6, "Intelligence": 7, "fun": 8

After the learning process:

(13) What is the optimal initial state vector π ? [6 marks]

(14) What is the probability of transitioning from state 5 to state 1? [6 marks]

C. Sentiment Analysis on Financial News: [55 marks]

We want to create a sentiment analysis model to classify some financial news into positive / negative sentences.

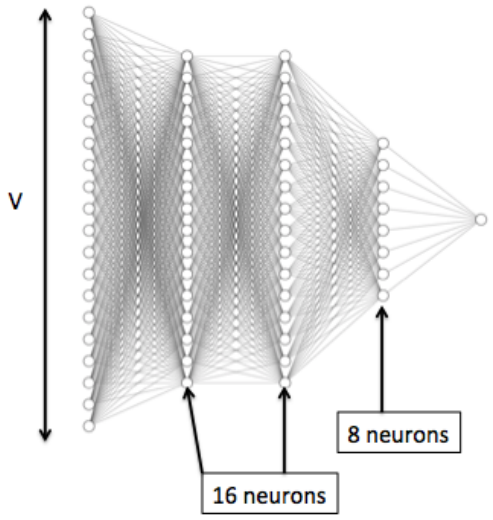
We use the following training dataset:

	Features				Target
Sentence 1	W1,1	...	...	W1,T	1
Sentence 2	W2,1	...	...	W2,T	0
...		...	...		
Sentence N	WN,1	...	...	WN,T	1

Let V be the vocabulary size.

C.1. A Deep Neural Network

We wish to use a Deep Neural Network with 3 hidden layers to perform the classification task.



Before feeding the data to the neural network, we first need to preprocess the sentences using a dictionary that maps the words into integers.

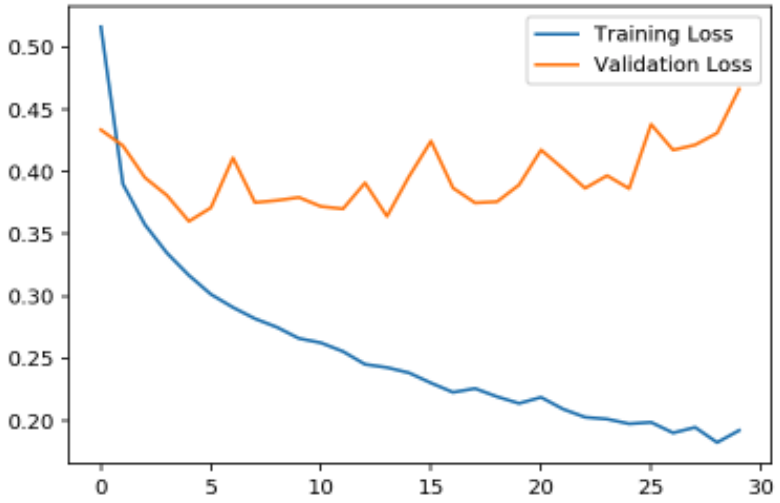
After transforming the sentences into sequences of integers, we decided to one hot encode each sequence of integers into a V – dimensional vector.

(15) What would be the shape of the tensor data after the preprocessing steps [4 marks]

(16) Which activation functions should we use? [6 marks]

(17) Deduce the loss function associated with this classification problem. [6 marks]

After fitting the data to the model, we end up with the following training and validation losses:



(18) Describe the problem highlighted by the validation loss? [6 marks]

(19) How would you solve it? [6 marks]

(20) Explain why the previous model will assign the same prediction to the sentences "What a great news in such a bad period" and "What a bad news in such a great period"? [6 marks]

(21) Describe the different steps of creating a generative classifier using HMMs to perform the same classification task [6 marks]

C.2. A Sequential Neural Network

We want to use a Sequential Neural Network with LSTM layers and pretrained D-dimensional word vectors to perform the same classification task.

(22) Describe such a model by specifying the different layers, the different hyperparameters for each layer and how the shape of the data is changing after each layer transformation. [15 marks]