Deep Learning for Artificial Intelligence (DLAI)
MET - ETSETB **MidTerm 7/11/2017**
50 min **Name:**

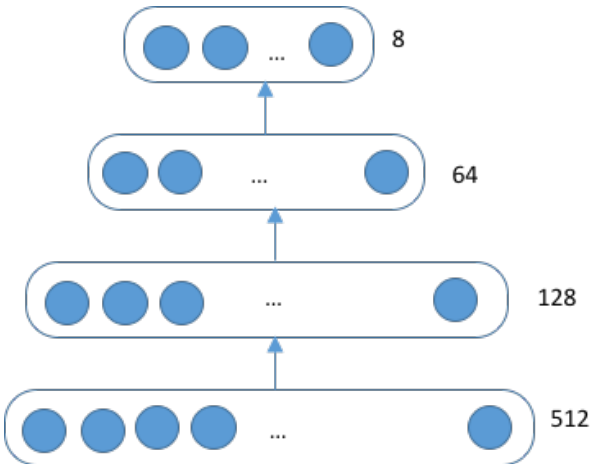
Question 1 Perceptron

Consider a binary classifier implemented with a single neuron modelled by two weights w_1=0.2 and w_2=0.8 and a bias b=-1.
Consider the activation function to be a sigmoid $f(x) = 1 / (1+e^{-x})$.

a) Draw a scheme of the model.
b) Compute the output of the logistic regressor for a given input x=[1,1].
c) Considering a classification threshold of y_{th}=0 (y_{th}>0 for class A, and y_{th}<0 for class B), which class would be predicted for the considered input x=[1,1] ?

Question 2 Multilayer Perceptron

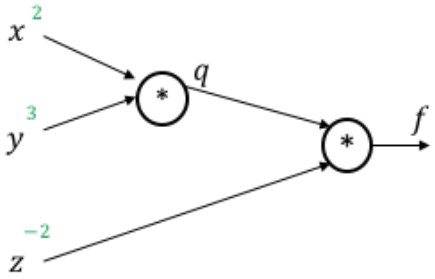
How many parameters do you need to compute in the following Multilayer Perceptron? (justify your answer, indicate the operation you have to carry out)



Question 3 Backpropagation

Compute $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ for the following Computational Graph

$f(x, y, z) = (xy)z$ (x=2, y=-3, z=-2)



Answer: (justify your answer)

a) $\frac{\partial f}{\partial x} =$

b) $\frac{\partial f}{\partial y} =$

c) $\frac{\partial f}{\partial z} =$

Question 4 Optimization

Consider the following types of *critical points* of a smooth function where the gradient is a zero vector: a *minima* is a point where all eigenvalues of the Hessian are positive, and a *saddle point* is point at which there is at least one positive and one negative eigenvalue. Now, answer the following questions

a) A numerical optimization technique to find the minima of a high dimensional function is more likely to encounter which of the two types of critical points? Give reason(s).

b) Is finding the global minimum of a neural network's training loss function important? Give reason(s).

c) Which one of the two escapes a saddle point more easily – gradient descent or stochastic gradient descent? Give reason(s).

Question 5 Convolutional Neural Networks

a) A 2D convolutional network has an input of size NxNxC (N= width, N= height, C channels). The first layer is a convolutional layer with K filters of size F, with stride 1 and padding P.

a1) Give a formula for the size of the feature maps

a2) Give a formula for the number of parameters

b) What is the use of 1x1 convolutions?

c) What is pooling, where is it used, and what is its purpose?

d) Explain how the size of a feature map between two convolutional layers can be reduced just like pooling but without using pooling.

Question 6 Loss functions

Assume a network that has been trained to classify images into 3 classes. Consider the following three vectors when evaluating the network on a test image input.

(a) $\begin{bmatrix}1\\0\\0\end{bmatrix}$ (b) $\begin{bmatrix}2.0\\1.0\\0.1\end{bmatrix}$ (c) $\begin{bmatrix}0.7\\0.2\\0.1\end{bmatrix}$ (d) 0.3567

Justify which vector correspond to (A) the scores (logits) of the network, (B) The softmax ouput, (C) the one-hot encoding of the labels and (D) the cross-entropy loss

Question 7 Transfer Learning and domain adaptation

In a NN, explain the concept of finetuning, the advantages it offers over a full re-train on the target task. Explain also when it can be applied.