Exam in Machine Learning and Deep Learning, Fall 2023

To be solved without any aids on Monday December 11th, 2023

Clearly explain your solution to each problem and carefully present all inter-mediary steps. You do not need to carry out any numerical calculations, that is, it is enough to just write the expressions and formulas without the calculations needed to produce the final numerical answer.

1. (20%)

   The following table contains training instances classified as $+$ or $-$ using two Boolean predictors a_1 and a_2.

Instance	Classification	a_1	a_2
1	$+$	T	T
2	$+$	T	T
3	$-$	T	F
4	$+$	F	F
5	$-$	F	T
6	$-$	F	T

   Given a discrete random variable X with two possible outcomes, which occur with probability p and q respectively, the entropy of X is formally defined as

   $$-p \log_2 p - q \log_2 q.$$

   (a) Write down the expression for the entropy of the classification column, that is, the entropy for $\{+, +, -, +, -, -\}$.

   (b) Explain how the root of a decision tree would be chosen for this dataset using entropy heuristics.

2. (20%)

   (a) The following table contains coordinates on a map and whether there is sea or land at each coordinate.

x coordinate	y coordinate	classification
3	3	land
1	3	sea
-1	0	land
2	1	land
-1	3	sea

   Construct a neural net that perfectly classifies this dataset. Remember to explain all steps needed to find the net.

   (b) One of the simplest useful optimization algorithms for neural nets is gradient descent. Consider the function $y = (x - 2)^2$. Use a learning rate of 0.1 to perform the first two gradient descent steps starting at $x_0 = 4$.

3. (20%) The following figure shows a convolutional neural net (CNN).
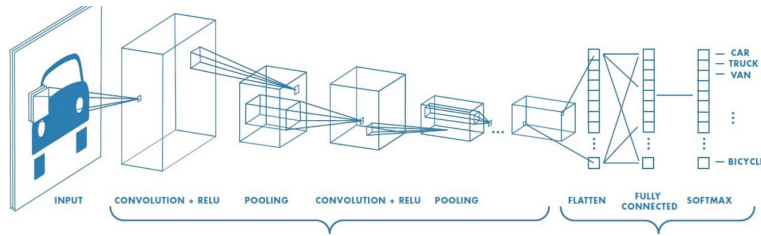
Explain all building blocks in this CNN as carefully as possible and also describe the overall architecture.

4. (20%) The following figure shows a long short term memory (LSTM) neuron.
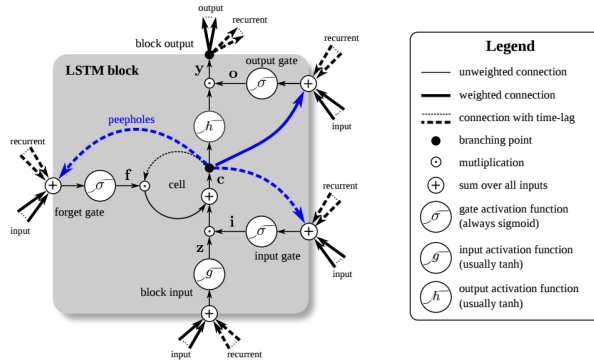
Explain all connections in this neuron and also explain how it works as thoroughly as possible,

5. (20%) Many large language models use Transformer layers. The most important part of a Transformer is full self attention. Write down the formulas for calculating attention weights using softmax attention in a Transformer. Also, describe how the output from full self attention is computed and why it is great for understanding and generating sentences in a human language.