Deep Learning topics:

Maximum Likelihood Estimation

Q:

Neural Networks

Q: Universal Approximation Theorem

Q: Explain if and why we have to introduce nonlinearity between 2 consecutive layers. Motivate your answer. For simplicity consider the case of fully connected (dense) layers.

Q: Explain why it is not a good idea to initialise all the weights of a Neural Network to zero. Consider both the cases in which the ReLU activation function is used, and the general case.

Q: Why is it convenient to use more that one hidden layer in neural networks? In other words, what is the advantage of a multi-layer neural network over a single-hidden-layer neural network?.

Backpropagation

Q: Calculate backpropagation for dense network, BPTT, RTRL

Regularization

Q: Explain what is dropout and how to implement it both in training and evaluation and why is that. Motivate your answer. Moreover, explain why it acts as regularization.

Q: Which techniques could be considered as a form of regularization?

Optimization

Q: What is the difference between backpropagation and (stochastic) gradient descent?

Q: Explain the main characteristics and the differences among: gradient descent, minibatch stochastic gradient descent, Adagrad, RMSprop and Adam.

Q: What are local minima and saddle point and why are they a problem for optimization algorithms?

Q: What are the main problems we face when optimising deep neural networks? Foe each one, explain why it is a problem for optimisation algorithms. Where applicable, explain how it is possible to avoid such problems.

CNN

Q:

Graph Neural Networks

Q:

Recurrent NN

Q 1: Give the definition of sequential transduction, explaining the concept of memory, causality, and recursive state representation. For each different type of sequential transduction we discussed in the lectures describe a network architecture able to implement it.

Q 2: When is sequential transduction casual?

Q 3: In the context of sequential transductions, give the definition of causality and discuss how this concept is implemented in Recurrent Neural Networks (RNN). Are all RNN architectures causal ?

- Back propagation through time
- Real time recurrent learning
- Gradient Vanishing/Exploding Problem
  Q: Question about Vanishing/exploging gradient, how to resolve the problem
  Q: Explain what a Reservoir Computing Network is. What are the main features and properties that such model owns?
- Cliping gradients
- Long Short-Term Memory Units. Gated Recurrent Units.
- Reservoir Computing. Echo State Networks and Liquid State Machines. Echo state property. Intrinsic Plasticity

Transformers

Q: Explain in detail the self-attention mechanism used in Transformers.

Autoencoders

Q: Define a Sparse Autoencoder. Explain the difference with respect to an undercomplete Autoencoder.

Q: The main feature of a Denoising Autoencoder:

Structured probabilistic models

Q: Kullback-Leibler divergence

Q1: What is a Monte Carlo Chain ? Explain why it is an important concept. Describe an algorithm for sampling from a Monte Carlo Chain.

Q: What is the difference between a probabilistic model represented by a directed graph and one represented by an undirected graph? Explain the pros and cons of each representation.

Q: What is the Differentiable Generator Network? Why is it useful? Give a simple example of a Differentiable Generator Network.

Q: Explain the meaning of ELBO, giving all technical details related to it. In addition, explain why ELBO is important.

Q: Training of a Restricted Boltzmann Machine is performed thanks to: