1. Project

- Some questions about the project implementation choices
- In the code, where can you find a specific function that's coming from theory? What is the mapping between theory statements about a certain function and the code implementing that function?
- Difference between RNN and LSTM (due to their presence in a project)
- Given a 3x3 NumPy array, extract the first two columns
    - Answer: matrix[:, :2]

2. Pre-Test

- Formula of Residual Connection
    - Answer: Simply $f(x) + x$
- Difference between Supervised and Unsupervised
- Very briefly, what is overfitting
- What is convolution and how do we use it in Deep Learning
- Why do we use the validation set?
    - Answer: To actually test our model on unseen data in order to see how well it generalizes.
- Difference between Regression and Classification
- Formula of the Binary Cross-Entropy Loss (BCE)
- What is ReLU and how would you write it in code
    - Answer: $\max(0, \alpha)$

3. Theory

3.1 Shallow and Deep Neural Networks

- Analytic form of a Shallow Neural Network
- Universal Approximation Theorem

3.2 Fitting and Backpropagation

- Update rule of Stochastic Gradient Descent (SGD)
- What's the difference between GD and SGD and what changes in their respective update rule?
- Given an MLP, find the number the parameters and the gradient w.r.t a certain weight

3.3 Measuring Performance and Regularization

- General overview of Regularization in Deep Learning
- Explain what is overfitting
- What is LeakyReLU function and why is it used?

3.4 CNNs and ResNets

- What is a Convolution?
- What is valid / same padding?
- CNNs uses and properties
- Residual Blocks: Definition & Formulas
- What's the role of Layer Normalization? Do you know any other type of Layer Normalization?

3.5 Self-Supervised Learning

- Self-Supervised Learning (SSL): How is the Dataset composed?
  - Answer: In SSL, the Dataset is composed only from samples x_i, since the labels are inferred directly from those. It's like as if the model learns to predict a part of the input from other parts of the input (such as recognizing structures etc.)
- Siamese Neural Networks
- What is Contrastive Learning? Apart from Self-Supervision, where is it used?
- Explain the Contrastive Loss and write the formula
- Explain the Triplet Loss
- In BYOL, does the Target Network perform backpropagation?

3.6 Transformers

- Optimization algorithm in RNNs
  - Answer: Backpropagation through time (+ explanation)
- Is Backpropagation through time (BPTT) expensive?
  - Answer: It's expensive, especially for long sequences, as the algorithm needs to remember and update parameters through multiple steps.
- What is one of the risks of BPTT?
  - Answer: Vanishing and exploding gradients. During the backward pass, gradients can become really small (vanishing) or very large (exploding). This makes difficult to update weights in earlier layers in an efficient way.
- Seq2Seq can be done using CNNs? How do you deal with variable length?
  - Answer: Seq2Seq could be deployed in CNNs and as for the variable length, one could exploit padding techniques. More technically, the default RNN Encoder gets replaced by a CNN which extract meaningful features, passed into a CNN/RNN Decoder which generates tokens (captions word-by-word) starting directly from those features.
- General overview of Attention
- What are the Self-Attention variables?
  - Answer: Query, Key, Value
- What is the hidden meaning behind $Q \cdot K^T$ in Self-Attention?
- How Q, K, V are used in a Transformer model? Why are they called like that?

3.7 Generative AI Foundations

- GAN: Architecture, Training, Losses and Nash Equilibrium
- GAN: Why do we need the Wasserstein GANs
- What's the problem with GANs and how to solve it
    - Answer: The major problem is the collapse of the Network. This can be stabilized through (for example) Wasserstein GNNs where we try to directly minimize the distance between the 2 distributions, discarding the possibility of involving the distributions in the loss as per normal GNN.
- VAE: Architecture
- VAE: Why do we need another Encoder function in VAE with different parameters ϕ?
- VAE: Why do we put ϵ? What is it needed for?
    - Answer: It's needed for the Reparameterization Trick (+ further explanation)

3.8 Graph Neural Networks

- GNN: What is Edge Classification?
- GNN: Inductive vs Transductive
    - Answer: Very simply, an Inductive GNN has multiple separated Training graphs, each one containing *Training + Validation* nodes. Along with that, it has one single Test graph. In a Transductive setting, the Training + Validation + Test nodes can be found in one unique linked big graph (It's really expensive though and not truly reasonable).
- Update rule of Graph Neural Networks (GNN)
- What are the various aggregation functions of a GNN?
    - Answer: Diagonal Enhancement, Mean Aggregation, Kipf Aggregation, Max Pooling Aggregation and finally Aggregation by Attention (**more on paragraph 2.3 – Aggregation in Chapter 10 of Summaries**).
- How does Attention work in GCN?
    - Answer: In this case, we calculate similarity scores between node pairs and normalize them through a SoftMax operation. After this, each node aggregates messages coming from their neighborhood (*weighted* by the previous scores), adds its features and updates its own representation. The main difference from the standard message passing relies on the fact that the neighbor's features are weighted by the similarity scores (attention).
- What's the disadvantage of using a GCN Spectrum based?
- What is a Graph Convolutional Network (GCN)?
- Message passing in GNN:
    - Answer: The core idea consists on updating node representation based on node neighborhoods. Each node creates a message based on its own features and its neighbor's features. After the creation, each single node aggregates these messages and combine with its own features in order to update its representation.

3.9 Geometric Deep Learning

- What is Geometric Deep Learning and explain me a transformation
    - Answer: A transformation could be Shift Equivariant / Invariant (explain their use in CNNs)

3.10 Model Compression

- General overview of Model Compression
- What is the Lottery Ticket Hypothesis?

3.11 Meta-Learning

- What is Meta-Learning? Explain an algorithm (for example Update rule of Optimization-Based)
- Explain a technique used in Metric-Based approach

3.12 Seminars

- What did you understand from the “Lecture on Learning in the Presence of Noisy Labels” part and if you know any modality of identifying corrupted labels
- What is Retrieval Augmented Generation (RAG)?