Discussion L3.1 (Generalisation and regularisation)

1) How would you detect underfitting and overfitting, how would they manifest themselves? What would you do to quickly handle these undesirable effects if you detected them?

2) In what scenarios (concerning data and neural network model characteristics) would you consider regularisation to be of critical importance?

3) What are the main components of the generalisation error? What do they tell us about the problem, data, neural network model etc.?

4) Does “more training data” approach always help with generalization?

5) How would you combine early stopping and model selection (hyperparameter selection) into a cross-validation scheme?

Discussion L3.2 (Generalisation and regularisation)

1) What is cross-validation particularly useful for? What are key drawbacks of this validation scheme?

2) What are extra challenges with regularisation using a penalty term in the objective function? How could you handle them?

3) What is the key difference between L1- and L2-norm (weight decay) regularisation? What are implications for the weight distribution? How can we interpret L1- and L2-norm regularisation in the Bayesian framework?

4) What is in your view the most appealing implication, consequence or effect of Bayesian regularisation?

5) How would you detect underfitting and overfitting, how would they manifest themselves? What would you do to quickly handle these undesirable effects if you detected them?

6) In what scenarios (concerning data and neural network model characteristics) would you consider regularisation to be of critical importance?

Discussion L3.3 (Ensembles)

1) In what situations would you consider using ensembles? How would you evaluate their usefulness, i.e. what would you compare with ensembles and what criteria would you rely on in your comparative analysis? Think about this from the perspective of a data scientist who has to solve a concrete problem and an ensemble is one of approaches to consider. In the end, you will have to defend your decision.

2) In what scenarios would a mixture-of-experts network approach be particularly applicable?

3) Let's say you have 15 MLP binary classifiers with 0.8 accuracy rate. If you combine them into a committee with a majority vote rule, what is the probability of a misclassification?

Discussion L3.4 (Case study)

1. How would you approach the following case study?

   You have data offered by a single manufacturer of steel parts:
   - Three different steel components (different steel types): 120 data points per one component type
   - Data sample consists of 52 measurements from the manufacturing process and has information about the tool wear (how much it deteriorated by manufacturing a single component)
   - Your task is to predict the tool wear or rather predict where the wear deteriorates below a certain fixed threshold
   - Ultimately, you would like to use your MLP based approach for steel components produced by other manufacturers

   How would you approach this problem, what are your key concerns from the generalization perspective? How do you deal with the data to maximise generalization capability of your method?