Mathematics of Machine Learning
Example Questions

Chapter 0 and Chapter 1

- What is machine learning?

- What types of machine learning are there and which one did we focus on?

- Explain the formal learning model!

- How do we evaluate and compare hypotheses?

- What is the (biased) ERM learning rule?

- What is the generalization error and how can it be decomposed?

Chapter 2

- What does PAC learnability mean?

- What is the No-Free-Lunch theorem and which conclusions can we draw from it?

- What is uniform convergence and how is it related to PAC learnability? (Can you proof that relation?)

- What is shattering and the VC dimension?

- What is the growth function and its relation to the VC dimension?

- What is the fundamental theorem of PAC learning theory?

- Give one example of a PAC-learnable class and a non-PAC-learnable class.

Chapter 3

- What are linear hypotheses and what is their geometric interpretation?

- What is the VC dimension of linear hypotheses?

- What linear classification methods do you know? Explain one of them in detail.

- What is the perceptron algorithm and what can you say about its convergence?

- What is the idea of hard vector machines?

- What is the idea of soft vector machines?

- Explain the "support vector" in SVM?

- What can you say about the sample complexity of SVM rules?

- What are kernels and what role do they play in machine learning?

- What is the kernel trick and what does the representer theorem state?

Chapter 4

- What is validation and the test error?

- Explain cross validation.

- Explain gradient descent. What do you know about its convergence?

- What are L-smooth and λ-strongly convex functions? Can you give examples?

- What is the stochastic gradient method?

- What are the advantages and disadvantages of the stochastic gradient method?

Chapter 5

- What is the architecture of a neural network?

- Describe the structure and working of a neural network? Give a formal definition of a FNN!

- Are neural networks PAC learnable? What do you know about their VC dimension?

- What are the univerisal approximation theorems for FNN?

- What is the approximation error for binary classification of shallow, arbitrarily wide FNN?

- Why is depth better than width?

- Give the main idea on how to train FNN.

- What are the advantages and disadvantages of FNN compared to the methods from Chapter 3?