

Oral Exam Questions

The following questions will be relevant for the oral exam of the Machine Learning class. Note that the questions should be considered as “openers”, i.e. subsequent questions that arise during the conversation in the exam are not included in this list. In the exam, usually 3 to 5 of the questions given here will be drawn by chance with the aim to provide a balanced coverage of topics.

- 0 Write the law of total probability, Bayes rule, and the definition of independence and conditional independence of two random variables.
- 1 Name and describe the categories of learning. Name two concrete examples for learning algorithms.
- 2 How does k -means clustering work?
- 3 What is the purpose of regression and what is the general principle?
- 4 Explain linear regression and what exactly is linear there.
- 5 What is a pseudoinverse matrix?
- 6 Name two types of basis functions and explain how they should be used.
- 7 Explain regularization in the context of regression. Why is this useful?
- 8 What is the probabilistic formulation of regression? What is the equivalent of maximum-likelihood and maximum-a-posteriori?
- 9 Name the different kinds of Probabilistic Graphical Models and draw two examples. What do they represent?
- 0 What is a perfect map between a graphical model and a joint probability distribution?
- 1 What is D-separation?
- 2 What is the definition of the Markov blanket and how can it be computed in directed and undirected graphical models?
- 3 What is the definition of a Hidden Markov Model?
- 4 How does the Viterbi algorithm work?

- ☒5 What is “explaining away” in the context of graphical models?
- ☒6 What is a Markov Random Field?
- ☒7 How does the expressiveness of directed and undirected graphical models compare?
- ☒8 What is the general idea of the inference on an undirected Markov Chain?
- ☒9 What is a factor graph?
- ☒0 What does the sum-product algorithm compute and how?
- ☒1 What is the principle of the AdaBoost algorithm?
- ☒2 Name the definition of a weak and a strong classifier. Give examples for each.
- ☒3 What other variants of AdaBoost exist?
- ☒4 How can AdaBoost be used for face detection?
- ☒5 Name 3 different activation functions used in Neural Networks. What is the advantage of ReLU over sigmoids?
- ☒6 What is the name of the training procedure in Neural Networks? How does it work?
- ☒7 Name the advantages of CNNs over fully connected NNs.
- ☒8 Sketch a perceptron for binary classification with n inputs.
- ☒9 What is a kernel function?
- ☒0 Name the Theorem of Mercer and explain the kernel trick.
- ☒1 How can kernelization be applied to k-means clustering?
- ☒2 What is the principle of kernel PCA?
- ☒3 What is a Support Vector Machine?
- ☒4 Name the definition of a Gaussian process. What specifies a GP? Draw an example.

- ⊞5 How can a function be sampled from a Gaussian Process?
- ⊞6 What is the predictive distribution using a Gaussian Process?
- ⊞7 What are the hyperparameters of the squared exponential kernel?
- ⊞8 What is Automatic Relevance Determination?
- ⊞9 What is the main problem in Gaussian Process classification in contrast to regression and how can it be solved?
- ⊞0 Name the definition of a Gaussian Mixture Model. What are the individual components and where can it be used?
- ⊞1 What are the differences and similarities between k-means and Expectation Maximization in Gaussian Mixture Models?
- ⊞2 What is the complete-data log-likelihood and how is it used in EM?
- ⊞3 What exactly is optimized in EM and how?
- ⊞4 Name examples for applications of the EM algorithm.
- ⊞5 What is variational inference?
- ⊞6 What is a conjugate prior?
- ⊞7 Name the definition and the properties of the Kullback-Leibler divergence. Where is it used?
- ⊞8 What are the main ideas of mean field and expectation propagation?
- ⊞9 What is a cavity distribution?
- ⊞0 Name and describe briefly 4 different sampling methods.
- ⊞1 How can a Particle Filter be used for Markov Localization?
- ⊞2 What is a stationary distribution in a Markov Chain?
- ⊞3 How does the Metropolis-Hastings algorithm work?
- ⊞4 What is Gibbs Sampling?

- ⑤ What is a Dirichlet Process and how can we construct one?
- ⑥ What is the idea of the Chinese Restaurant Process?
- ⑦ What is the principle of Affinity Propagation?
- ⑧ How does Spectral Clustering work?
- ⑨ What is agglomerative clustering?