Theoretical minimum

1. What is i.i.d. data?
2. What is the likelihood? Where is Maximum Likelihood Estimation (MLE) usually used?
3. Supervised learning problem statement
4. How does a Naive Bayesian Classifier work? Why is it naive?
5. Linear regression model for MSE minimization problem. Write down the formula and the derivative of the loss function w.r.t. weights.
6. Write down a gradient descent step for MSE loss in linear regression. How to use it with huge datasets?
7. What is cross-validation? How does the number of folds affect the validation?
8. What is overfitting and underfitting? How to detect them?
9. What is the difference between parameters and hyperparameters? Provide an example for linear models and decision trees.
10. What is a regularization? What is the difference between L1 and L2 regularization in linear models? Is it the only way to constrain the solution?
11. Does L2 regularization regularize the bias term (w_0 or b)? Why?
12. Why is it a good idea to normalize data before applying a linear model?
13. Provide a linear classification problem statement. What is a margin?
14. What are precision and recall? How to use them to measure the model quality?
15. Assume the dataset for binary classification is imbalanced, so 95% of data belong to the first class. How to adjust the classification quality measures, to work with such data?
16. Logistic loss function. How is it related to Maximum likelihood estimation?
17. Describe the greedy optimization algorithm for the decision tree.
18. Why can an unconstrained decision tree achieve zero error on the training set with all unique objects?
19. How to assign a class label for the object in the tree leaf in classification?
20. How to assign a class label for the object in the tree leaf in regression? Does it depend on the information criteria?
21. What is bagging?
22. What is Random Forest? How does it differ from Bagging over decision trees?
23. How are base algorithms being trained in gradient boosting?
24. How does backpropagation work in neural networks? What will be vector by vector derivative?
25. How does the Convolutional layer work? What is the convolution operation?

26. Why fully connected (dense) networks are not the best choice to work with image data? Why do CNNs perform better?
27. How does basic RNN (Vanilla RNN) work?
28. How does dropout work?
29. How do dropout and batch normalization change their behaviour on the inference stage?

General questions for self-check

1. Machine Learning problem statement. Regression, Classification, examples.
2. How to measure quality in classification: accuracy, balanced accuracy, precision, recall, f1-score, ROC-AUC, multiclass extensions.
3. How to measure quality in regression: MSE, MAE, R2.
4. Maximum likelihood estimation, how is it related to regression and classification
5. Naive bayesian classifier, how does it work
6. K-nearest neighbours classifier, how does it work
7. Linear regression. Problem statement for the MSE loss function case. Analytical solution. Gauss-Markov theorem. Gradient approach in linear regression.
8. Regularization in linear models: L1 и L2, their properties. Probabilistic interpretation.
9. Logistic regression. Equivalence of MLE approach and logistic loss minimization.
10. Multiclass classification. One-vs-one, one-vs-all, their properties.
11. Support vector machine. Optimization problem for SVM. Kernel trick. Kernel properties.
12. Principal component analysis. Relations to SVD. Eckart-Young theorem. How to apply PCA in practice.
13. Train, validation and test stages of model development.Overfitting problem, ways to detect it.
14. Validation strategies. Cross validation. Data leaks.
15. Bias-variance tradeoff.
16. Decision tree construction procedure.
17. Information criteria. Entropy criteria, Giny impurity.
18. Ensembling methods. Bootstrap. Bagging.
19. Random Forest, Random subspace method.
20. Boosting and gradient boosting. Main idea, gradient derivation.
21. Matrix calculus and matrix derivatives. How to get the derivative of matrix/dot product, e.g.: aTx, Ax.
22. Backpropagation, chain rule.
23. Neural network concept. Fully-Connected layer (FC). Logistic regression as simple NN.
24. Losses for NNs: logistic loss, cross-entropy.

25. Activation functions, their impact on the network, computational complexity. Softmax and LogSoftmax activations, numerical stability.
26. Optimization methods in Deep Learning. Gradient descent, SGD, it upgrades: Momentum, RMSProp, Adam.
27. Regularization in Deep Learning: Dropout, Batch Normalization. Differences in training and evaluation stages.
28. Vanilla Recursive NN cell. Backpropagation through RNN. Vanishing gradient problem. Potential solutions.
29. Matrix convolution. Convolutional layer, backpropagation through it. Hyperparameters of Convs. 1x1 convolutions, comparison to FC layers. Max/Average Pooling.