Deep Learning

February 2020

1. In Machine Learning, we usually follow the approach known as Empirical Risk minimization:

$$R_{emp}(h) = \frac{1}{N} \sum_{n=1}^{n} L(h(x_n), y_n)$$

   Why is $\hat{y}$, the hypothesis that minimizes the empirical risk on the training set $\{(x_n, y_n)\}_{n=1}^{N}$ not necessarily the optimal solution?

   - Chapter: and slides:
   - the formular is not the important thing here, it looks like if it is not relevant to solve the question.
   - by the empirical risk the loss is relvant

2. Suppose you have completed a first training run of a deep neural network and observe the following loss curves for training and validation set. What is your diagnosis and what steps could you take to improve the model? GRAPH

3. Suppose with have a CNN with the following layers:

   (a) Input images: RGB with 12x12 pixels
   (b) 4 kernels of size 4x4, stride 2, no zero padding (valid-padding)
   (c) 8 kernels of size 3x3, stride 1, no zero padding (valid-padding)
   (d) 16 kernels of size 3x3, stride 1, no zero padding (valid-padding)

   - What is the receptive field size in pixels of a unit in the second layer?
   - What is the dim of the output?
   - How many parameters does the network have? Let's assume there is no batch normalization and we don't use biases in the convolutional layers. You can write down the formula without calculating the final result.

4. What is the purpose of residuals connections in ResNets and derivatives?

5. Which of the following two architectures is more suitable for applications where the computational resources should be kept a minimum?

   ☐ EfficientNet
   ☐ MobileNet
   ☐ ResNet

6. State an example of a sequence-to-sequence task in natural language processing.

7. What is the training objective of the discriminator in a GAN?

8. Which of the following generative models allows us to compute the likelihood of an image under the model?

   ☐ Variational Autoencoder
   ☐ GAN
   ☐ PixelCNN

9. State three important differences of deep learning practices in the 1990ies and today.

10. A feedforward neural network with finite width and ReLU activation function can approximate any continuous function to arbitrary precision.

    ☐ correct

    ☐ wrong

11. Over which dimensions of the Image x Channels x Height x Width tensor is the normalization performed in layer normalization?

12. Which of the following statements about Dropout (p=0.5) are correct (multiple answers possible)?

    ☐ During training 50% of the weights are set to zero in each iteration.

    ☐ 50% of the units are removed from the network at random.

    ☐ During inference, activations have to be scaled by a factor of 2.

    ☐ Dropout can be thought of as training an ensemble of models.

    ☐ The randomness introduced by Dropout during training hurts model performance.

13. Under what conditions can a perceptron learn to classify a training set of paires perfectly?

    ☐ The two classes are seperable by ab arbitrary decision bound

    ☐ The two classes are linearly seperable

    ☐ Perfect classification is not possible

14. Suppose we want to model the queue length at a grocery store given some predictores $x \in R^d$. We assume that the number of people in the queue ($k \in \{0, 1, 2, ...\}$) follows a Poisson distribution with rate λ:

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

    We model the rate λ using a deep neural network, i.e. $\lambda = f(x_i\theta)$ with parameter θ. What would be the loss function if we choose to minimize the negative log-likelihood of the observations are independent?

15. What diagnostic function does overfitting the neural network on a single minibatch of training data serve?

16. In a convolutional neural network, the number of parameters is independent of the image size

    ☐ Correct

    ☐ Wrong

17. How does a depthwise seperable convolution differ from regular convolution?

18. Name at least two commonly used mechanisms for regularizing neural network training.

19. For which of the following normalization schemes is the following statement correct?
    "_____ behaves differently during training and inference time."

    ☐ Batch normalization

    ☐ Layer normalization

    ☐ Group normalization

    ☐ Instance normalization

20. A friend asks you for advice on a project that involves recognization different car models. She has collected a labeled dataset containing 10 images for each of 50 different car models. Which of the following approaches would you recommand if classification accurancy is the primary objective? (pick one)

    ☐ Train a ResNet-18 from scratch

    ☐ Use a ResNet-18 pre-trained on ImageNet, replace and train only classification layer

    ☐ Use logistic regression

    ☐ Search for a small CNN architecture that performs well

    ☐ Use a ResNet-152 pre-trained on ImageNet, replace and train only classification layer