Question 1: Short Questions

(a) We use a support vector machine (SVM) with a soft-margin SVM to perform a classification task. Which of the following types of samples will have zero slack variables ξ_i?

(i) All support vectors

(ii) All correctly classified samples

(iii) All misclassified samples

(iv) All samples lying within the margin

(b) Explain in which of the following cases the risk of overfitting a network decreases.

(i) Regularizing the weights

(ii) Increasing the number of the hidden layers

(iii) Using dropout to train a deep neural network

(iv) Getting additional training data that are very similar to the training data that have been seen before

(c) You have a neural network with two inputs $x_1 = 2$, $x_2 = 2$, connected to the output with two weights $w_1 = 0.5$ and $w_2 = -0.2$. The bias term of the input is $b_1 = 0.1$. You use three different activation functions and get the following output for each function $\{0.7, 0.67, 1\}$. Which type of activation function was used for each of the three outputs?

(i) (linear, indicator/step, sigmoid)

(ii) (ReLU, sigmoid, indicator/step)

(iii) (ReLU, indicator/step, sigmoid)

(iv) (indicator/step, sigmoid, linear)

(d) Please select the correct answer(s).

(i) It is possible to successfully train a network by initializing all the weights to zero

(ii) It is not possible to successfully train a network by initializing all the weights to zero

(iii) It is possible to successfully train a network by initializing all the biases to zero

(e) The number of nodes in the input layer is 10 and the hidden layer is 5. The maximum number of connections from the input layer to the hidden layer is:

(i) 50

(ii) 10

(iii) 5

(iv) 55

(f) We perform a convolution operation to an input image of size 28×28 using a kernel/filter of size 7×7 with a stride of 1. What will be the size of the resulting convoluted matrix if we assume that there is not zero-padding at the boundaries of the image?

(i) 22×22

(ii) 28×28

(iii) 21×21

(iv) 7×7

(g) Which of following activation function is the most suitable at output layer of a neural network to classify an image in a binary classification task?

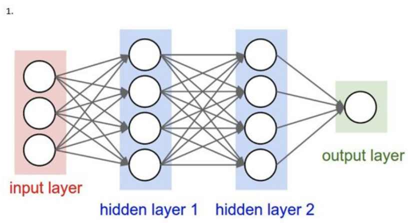
(i) Sigmoid

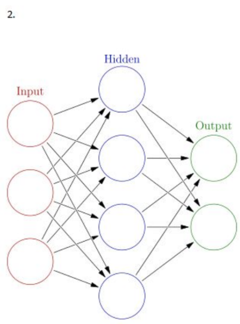
(ii) ReLU

(iii) Linear

(iv) None of the above

(h) For which type of activation function in the nodes of the hidden layers would Architecture 1 be equivalent to Architecture 2?





(i) Sigmoid

(ii) Hyperbolic tangent

(iii) Linear

(iv) ReLU

Question 2: Maximum likelihood estimation

The voters in a given town arrive at the place of voting according to a Poisson process of rate λ voters per hour, where $\lambda = 1 - (\theta t - 1)^2$ and $t = 1, \ldots, 12$. Using the Poisson distribution, we can express the probability of x voters coming to the poll within each hour using the following equation $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$. We further assume N samples $\mathcal{X} = \{(t_1, x_1), \ldots, (t_N, x_N)$ that represented the number of voters x_n that came to the poll at time t_n.

(a) Compute the likelihood of sample (t_n, x_n).

(b) Compute the likelihood of all samples $l(\mathcal{X})$.

(c) Compute the log-likelihood of all samples $logl(\mathcal{X})$.

(d) Describe how you would find the maximum likelihood estimate of θ given the samples $\mathcal{X}$.