Unit 8 - Week 5 - Approximation of functions, CNN and Cover's Theorem

Assignment 05

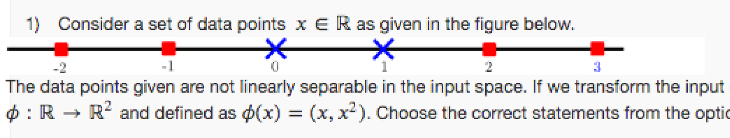
The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2019-09-04, 23:59 IST.

Instructions:

1. **Attempt all questions.**
2. **Submission deadline**: 4th September 2019 23:59 IST
3. **Solutions to be posted**: 5th September 2019
4. Older browsers might show **unnecessary vertical bars** at the end of math equations.

1) Consider a set of data points $x \in \mathbb{R}$ as given in the figure below. *2 points*



The data points given are not linearly separable in the input space. If we transform the input data into a higher dimension feature space using $\phi(.)$ function such that $\phi : \mathbb{R} \rightarrow \mathbb{R}^2$ and defined as $\phi(x) = (x, x^2)$. Choose the correct statements from the options.

- The feature space is not linearly separable.
- The feature space is linearly separable.
- There exists only a single hyperplane which can classify the data points in feature space.
- Multiple hyperplanes exists which can classify the data points in feature space.

No, the answer is incorrect.
Score: 0
Accepted Answers:
The feature space is linearly separable.
Multiple hyperplanes exists which can classify the data points in feature space.

2) Consider a convolutional neural network with K masks of size $(N \times N)$ in the convolutional layer. The input to the network is a grayscale image of size $(W \times H)$, i.e., depth is 1. The convolutional layer is followed by a max-pooling layer of size (2×2). What is the dimension of the output feature from the max-pooling layer? *2 points*

- $\left(\frac{W-N+K}{2} \times \frac{H-N+K}{2} \times K\right)$
- $\left(\frac{W-N+K}{2} \times \frac{H-N+K}{2} \times 2\right)$
- $\left(\frac{W-N+1}{2} \times \frac{H-N+1}{2} \times K\right)$
- $\left(\frac{W-N+1}{2} \times \frac{H-N+1}{2} \times \frac{K}{2}\right)$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\left(\frac{W-N+1}{2} \times \frac{H-N+1}{2} \times K\right)$

3) In K-fold cross-validation we randomly partition the training data into K sets of equal size and run the learning algorithm K times. Each time, the model is trained on the K-1 sets and tested on one set. If e_i is the mean square error by holding out the i^{th} set as the testing set, then an estimate of the true error is given by *2 points*

- $\sum_{i=1}^{K} ie_i$
- $\sum_{i=1}^{K} Ke_i$
- $\sum_{i=1}^{K} e_i$
- $\frac{1}{K} \sum_{i=1}^{K} e_i$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$\frac{1}{K} \sum_{i=1}^{K} e_i$

4) Choose the correct statement(s): *2 points*

- The K-fold cross-validation can be used to select tuning parameters for more complex models.
- The Leave-one-out-cross-validation is a better approach for training small datasets.
- Cross-validation approach consumes lesser time than the normal training approach.
- Cross-validation requires knowledge of the data distribution a priori.

No, the answer is incorrect.
Score: 0
Accepted Answers:
The K-fold cross-validation can be used to select tuning parameters for more complex models.
The Leave-one-out-cross-validation is a better approach for training small datasets.

5) (True/False) Back-propagation cannot be applied to pooling layers. *1 point*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False

6) Many learning algorithm can be viewed from an optimization framework. For example, in multi-layer perceptron we minimize the error between the output and the desired signal with respect to the network parameters. One important optimization technique is finding the Newton direction which requires computation of Hessian of the cost function. Let the cost function $f(x, y, z)$ be $f(x, y, z) = x^3 z + yz^2$ where x, y, z are the network parameters. Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$ is given by

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2}$$

The Frobenius norm of Hessian of $f(x, y, z)$ at $(1, 2, 3)$ is

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Range) 20,21

2.5 points

7) As discussed in class, Cover's theorem states that "A complex pattern-classification problem, cast in a high-dimensional space nonlinearly, is more likely to be linearly separable than in a low-dimensional space, provided that the space is not densely populated".

Consider a two bit XOR problem. Suppose we lift these data points to a higher dimension, we should be able to linearly separate the two classes. Therefore, consider applying the following transformation on the data points using four Gaussian hidden functions as follows

$\phi(\bar{x}_i, \bar{t}_j) = \exp\left(-\frac{\|\bar{x}_i - \bar{t}_j\|}{2\sigma^2}\right)$

where $\{\bar{x}_i\}_{i=1}^{4} = \{(0,0)^T, (0,1)^T, (1,0)^T, (1,1)^T\}$ are the inputs and $\{\bar{t}_i\}_{i=1}^{4} = \{(0,0)^T, (0,1)^T, (1,0)^T, (1,1)^T\}$ are the parameters corresponding to each Gaussian hidden function with a common $\sigma = 1$. Now, define a matrix $\Phi = \left[\phi(\bar{x}_i, \bar{t}_j)\right]_{i,j}$ then, determinant of Φ is

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Range) 0.1900,0.2100

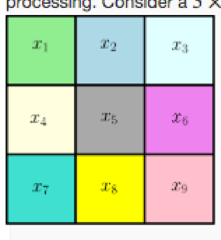
2 points

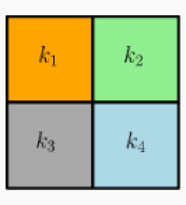
8) (True/False): The features extracted by the kernels in the convolution layer are rotation and scale invariant. *0.5 points*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False

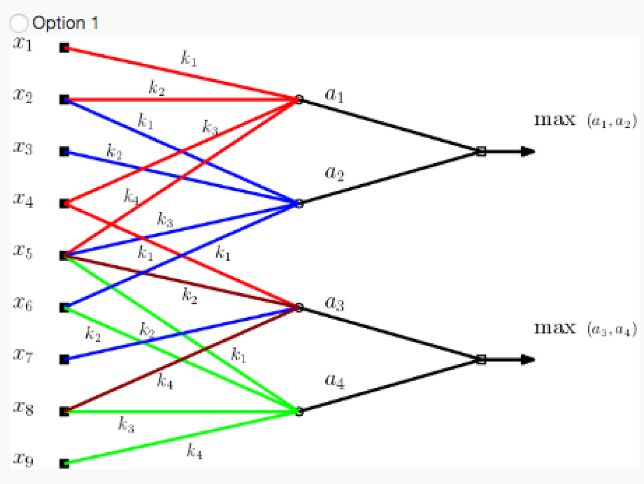
9) Consider the architecture of a convolution neural network over images using weight sharing i.e., the same filter or kernel is applied to various regions of an input image to extract the local features through that kernel before subsequent processing. Consider a 3×3 image and a 2×2 kernel as shown below. *3 points*



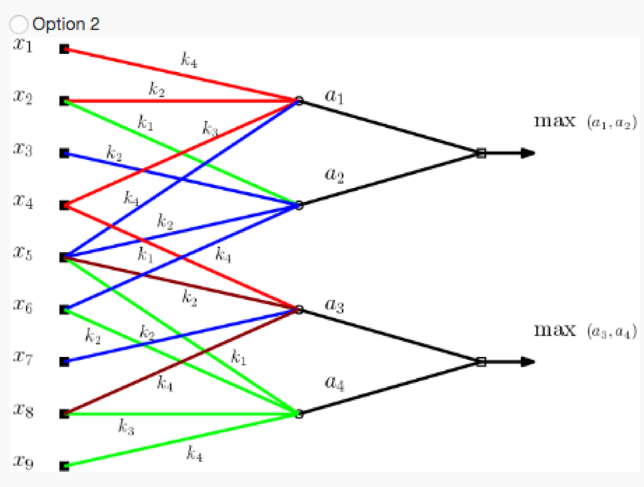


Suppose we want a 2×1 vector after max-pooling. By rastering the image as 1D vector and connecting the input nodes (in rastered form) to the convolution layer and subsequently to the pooling layer, the idea of weight sharing is clearly illustrated by

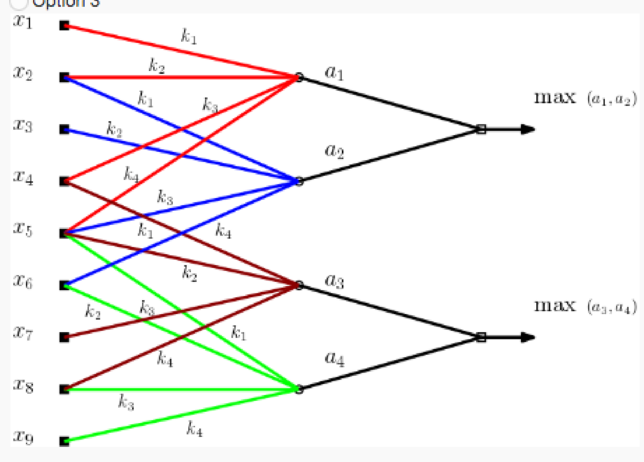
- Option 1



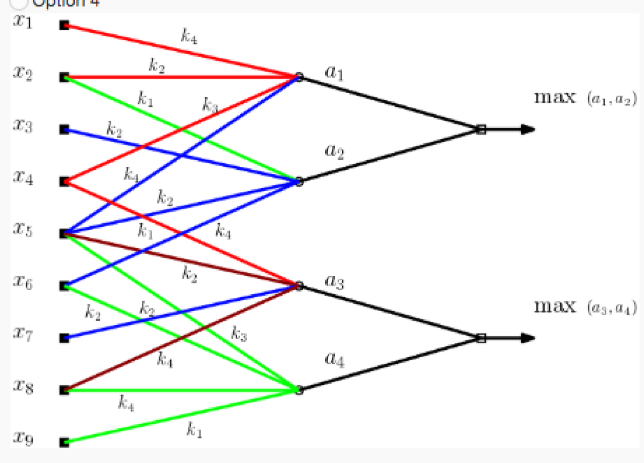
- Option 2



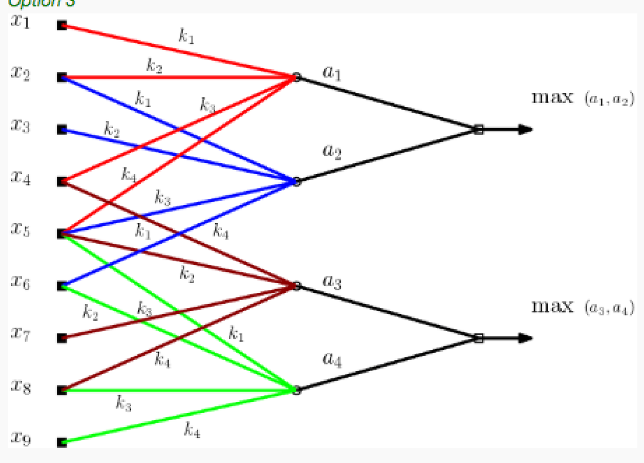
- Option 3



- Option 4



No, the answer is incorrect.
Score: 0
Accepted Answers:
Option 3



10) (True/False): Let $x \in [-5, 5]$ and define $f(x) = x^2$. Consider a network with a single hidden layer with arbitrarily large number of hidden neurons. The training of the network to learn the inverse mapping with input as $f(x)$ and output as x for all x converges in finite steps. *2 points*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False

11) (True/False) If the number of hidden layers are increased in a network, the classification error always decreases. *1 point*

- True
- False

No, the answer is incorrect.
Score: 0
Accepted Answers:
False