Unit 4 - Week 1

Course outline

How to access the portal

Matlab and Learning Modules

Pre-Requisite assignment

Week 1

- Introduction to the Course History of Artificial Intelligence
- Overview of Machine Learning
- Why Linear Algebra ? Scalars, Vectors, Tensors
- Basic Operations
- Norms
- Linear Combinations Span Linear Independence
- Matrix Operations Special Matrices Matrix Decompositions
- Quiz : Assignment 1
- Week - 1 Feedback Form

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

Week 12

Assignment 1

The due date for submitting this assignment has passed.

Assignment submitted on 2019-01-27, 01:31 IST

Due on 2019-02-13, 23:59 IST.

For decimal answers, provide answers to 4 decimal places (rounded off on the last decimal)

A (colour) 128x128 image is input into an algorithm which outputs a (colour) 16x16 image representing some important portions of the original image. For example, the input could be the image of a lung and the output could be a suspicious tumorlike portion, etc. (We encourage you to think of other examples, as an exercise). Answer questions 1 to 3 for this

1) If the input is turned into a vector x, its length would be?

```
49152
```

Yes, the answer is correct.
Score: 1

Accepted Answers:
(Type: Numeric) 49152

1 point

2) If the output is turned into a vector y, its length would be?

```
768
```

Yes, the answer is correct.
Score: 1

Accepted Answers:
(Type: Numeric) 768

1 point

3) If we write a model of y as y = Ax + b. Then, the total number of elements in the matrix A is?

```
37748736
```

Yes, the answer is correct.
Score: 1

Accepted Answers:
(Type: Numeric) 37748736

1 point

4) Let $x = \begin{bmatrix} -10 & 2 & 4 & 8 & 9 \end{bmatrix}^T$. Then, which of these is the greatest?   *1 point*

- ⦿ L_1 norm
- ○ L_2 norm
- ○ L_3 norm
- ○ L_∞ norm

Yes, the answer is correct.
Score: 1

Accepted Answers:
L_1 norm

Questions 5-10 refer to the same matrix W

Note : Some of the following questions might require you to write short programs in order to answer (or else it would require doing extremely long computations by hand). We recommend using MATLAB for simplicity, but any

programming environment will do.

Consider the following matrix $W = \begin{bmatrix} 1 & 3 & 2 & 4 & 6 \\ 3 & 2 & 7 & 8 & 7 \\ 2 & 7 & 3 & 7 & 8 \\ 4 & 8 & 7 & 4 & 9 \\ 6 & 7 & 8 & 9 & 5 \end{bmatrix}$. Answer the following questions.

5) What is the maximum eigenvalue of W?

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 28.76,28.78

1 point

6) What is the square root of the maximum eigenvalue of W^2?

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 28.76,28.78

1 point

7) What is the maximum singular value of W?

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 28.76,28.78

1 point

8) Which of the following is true? Choose all the correct answers   *1 point*

☐
$\lambda(M) = \sqrt{(\lambda(M^2))} = svd(M)$ for any real matrix M

☐
$\lambda(M) = \sqrt{(\lambda(M^2))} = svd(M)$ for any real, symmetric matrix M

☐
$\sqrt{(\lambda(M^2))}$ is always real for any real matrix M

☐
$\sqrt{(\lambda(M^2))}$ is always real for any real, symmetric matrix M

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\lambda(M) = \sqrt{(\lambda(M^2))} = svd(M)$ *for any real, symmetric matrix M*
$\sqrt{(\lambda(M^2))}$ *is always real for any real, symmetric matrix M*

9) Let $y_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^T$ and $b = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix}^T$. Let $y_n = Wy_{n-1} + b$. Then, $\|y_1\|_2$ is?

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 8.54,8.55

1 point

10) $\|y_2\|_2 / \|y_0\|_2$ is = ?

No, the answer is incorrect.
Score: 0

Accepted Answers:

(Type: Range) 234.86,234.87

1 point

Previous Page   End

Announcements   **Course**   Ask a Question   Progress   FAQ

Unit 5 - Week 2

Course outline

How to access the portal

Matlab and Learning Modules

Pre-Requisite assignment

Week 1

Week 2

○ Introduction to Probability Theory Discrete and Continuous Random Variables

○ Conditional, Joint, Marginal Probabilities Sum Rule and Product Rule Bayes' Theorem

○ Bayes' Theorem - Simple Examples

○ Independence Conditional Independence Chain Rule Of Probability

○ Expectation

○ Variance Covariance

○ Some Relations for Expectation and Covariance (Slightly Advanced)

○ Quiz : Assignment 2

○ Week 2 Feedback Form

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Assignment 2

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

Due on 2019-02-13, 23:59 IST.

1) The probability that the sum of the values of 2 die when thrown is equal to 11 is: *1 point*

- ○ 1/18
- ○ 1/36
- ○ 1/12
- ○ 1/9

No, the answer is incorrect.
Score: 0

Accepted Answers:
1/18

2) The probability that an ace is drawn on the second draw from a well shuffled pack of cards given that the first one was an ace is: *1 point*

- ○ 3/51
- ○ 4/51
- ○ 4/52
- ○ 3/52

No, the answer is incorrect.
Score: 0

Accepted Answers:
3/51

3) A family has two children. Given that one of the children is a boy, what is the probability that both children are boys?   *1 point*

- ○ $\frac{1}{2}$
- ○ $\frac{1}{4}$
- ○ $\frac{1}{3}$
- ○ $\frac{3}{4}$

No, the answer is incorrect.
Score: 0

Accepted Answers:
$\frac{1}{3}$

4) Which of the following statements is true?   *1 point*

- ○ Independent events must be mutually exclusive.
- ○ The sum of probabilities of mutually exclusive events must be 1.
- ○ The sum of probabilities of mutually exclusive and collectively exhaustive events must be 1.
- ○ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
The sum of probabilities of mutually exclusive and collectively exhaustive events must be 1.

5) If the random variable X follows the below distribution, what is the value of c?   *1 point*

[   ]

- ○ 4
- ○ 3
- ○ 2

○ 1

No, the answer is incorrect.
Score: 0

Accepted Answers:
4

6) Which of the following statements is true with regards to the probability distribution *1 point*
function f(x) of a random variable X?

○ f(x) must be less than 1 for all values of x
○ f(x) must be non-negative for all values of x
○ f(x) cannot exist for negative values of x
○ All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
f(x) must be non-negative for all values of x

7) An image is represented as a vector **x**. We wish to classify the image in one of 3 classes -- a *1 point*
cat, a dog or neither.

The classification output is represented as a vector **y** as follows. If it is a cat, then **y** =[1 0 0], if it
is a dog then **y** =[0 1 0] and if it is neither, then **y** =[0 0 1].
Someone creates an algorithm that takes in as input the image and output a probability vector **h**
-where each element gives the respective probability. For example, if **h** = [0.7 0.2 0.1], it means that
the given image has a probability of 0.7 that it is a cat, 0.2 that it is a dog and 0.1 that it is neither.
Which of the following statements is true? (Mark all that are correct)

☐
☐
☐

☐

No, the answer is incorrect.
Score: 0

Accepted Answers:

8) If a fair coin is tossed 4 times, what is the expected number of heads? *1 point*

○ 1
○ 2
○ 3
○ 4

No, the answer is incorrect.
Score: 0

Accepted Answers:
2

9) Given two random variables X and Y, which of the following equations hold true? *1 point*

☐ E[XY] = E[X]E[Y]
☐ var(X) + var(Y) = var(X+Y)
☐ If X and Y are independent, the covariance of X and Y is zero
☐ E[X+Y] = E[X] + E[Y]

No, the answer is incorrect.
Score: 0

Accepted Answers:
If X and Y are independent, the covariance of X and Y is zero
E[X+Y] = E[X] + E[Y]

10)What is the expected value of the random variable X with probability distribution function *1 point*
given

- ○ ½
- ○ ⅔
- ○ 2/9
- ○ 3/2

No, the answer is incorrect.
Score: 0
Accepted Answers:
3/2

Previous Page

End

- ○ ½
- ○ ⅔
- ○ 2/9
- ○ 3/2

Courses » Machine Learning for Engineering and Science Applications

Announcements   **Course**   Ask a Question   Progress   FAQ

Unit 6 - Week 3

Course outline

**How to access the
portal**

**Matlab and
Learning Modules**

**Pre-Requisite
assignment**

Week 1

Week 2

Week 3

○ Machine
Representation of
Numbers, Overflow,
Underflow,
Condition Number

○ Derivatives,Gradient,Hessian,
Series

○ Matrix Calculus
(Slightly Advanced)

○ Optimization – 1
Unconstrained
Optimization

○ Introduction to
Constrained
Optimization

○ Introduction to
Numerical
Optimization
Gradient Descent -
1

○ Gradient Descent –
2 Proof of Steepest
Descent Numerical
Gradient
Calculation
Stopping Criteria

○ Introduction to
Packages

○ Quiz : Assignment
3

○ Week - 3 Feedback
Form

○ ERRATA

Week 4

Week 5

Week 6

Week 7

Assignment 3

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

**Due on 2019-02-20, 23:59
IST.**

All questions may be answered by by writing a program in any language of your choice. Please
give answers to 4 decimal places of accuracy (round off the final decimal place). The solutions
assume that you have used double precision in your computation (which is the default in
MATLAB)

NOTE : All the norms asked here are 2-norms. In practice, you need to calculate the
components of the vector in order to calculate the norm.

1) ▫

▫

**No, the answer is incorrect.
Score: 0**

Accepted Answers:
(Type: Numeric) -20

1 point

2)

**No, the answer is incorrect.
Score: 0**

Accepted Answers:
(Type: Numeric) -2

1 point

3) ▫

**No, the answer is incorrect.
Score: 0**

Accepted Answers:
(Type: Numeric) 3

1 point

4)

**No, the answer is incorrect.
Score: 0**

Accepted Answers:
(Type: Numeric) -1.8208

1 point

5)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 3.0512

1 point

6)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 0.1823

1 point

Based on the data, Answer the from Question 7-10

7)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 2

1 point

8)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 1.0506

1 point

9)

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 0.7922

1 point

10)

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Range) 0.008,0.009

1 point

Previous Page

End

Unit 7 - Week 4

Register for Certification exam

Course outline

Assignment 4

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

Due on 2019-02-27, 23:59 IST.

1)   *1 point*

○ ▫
○ ▫
○ ▫
○ ▫

No, the answer is incorrect.
Score: 0

Accepted Answers:
▫

2)   *1 point*

○ ▫
○ ▫
○ ▫
○ ▫

No, the answer is incorrect.
Score: 0

Accepted Answers:
▫

3) Which of the following statements are True? Check all that apply: *1 point*

☐ If a learning algorithm is suffering from high bias, only adding more training examples may **not** improve the test error significantly.

☐ A model with more parameters is more prone to overfitting and typically has a higher variance.

☐ When debugging learning algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem.

☐ Increasing degree of the polynomial in curve fitting will increase the bias in the model

No, the answer is incorrect.
Score: 0

Accepted Answers:
If a learning algorithm is suffering from high bias, only adding more training examples may **not** *improve the test error significantly.*
A model with more parameters is more prone to overfitting and typically has a higher variance.
When debugging learning algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem.

4) The figure below shows the plot of the learning curves of a learning algorithm. It is found that it has an   *1 point*
unacceptably high error on the test set. What is the algorithm suffering?

○ High Variance
○ High Bias
○ High Variance and Low bias
○ None

No, the answer is incorrect.
Score: 0

Accepted Answers:
High Bias

5) Suppose you have implemented a regularized linear regression model. You observe that on the held out **1 point** testing set, the model makes unacceptably large errors with its predictions. However, you observe that the model performs well (has a low error) on the training set. Which of the following steps can be incorporated to lower the error on testing dataset. Select all that apply.

☐ Try using a smaller set of the features
☐ Try decreasing the regularization parameter λ
☐ Get more training examples
☐ Use fewer training examples

No, the answer is incorrect.
Score: 0

Accepted Answers:
Try using a smaller set of the features
Get more training examples

6) Suppose you have implemented a regularized linear regression model. You observe that on the held out **1 point** testing set, the model makes unacceptably large errors with its predictions. Furthermore, you observe that the model performs **poorly** on the training set. Which of the following steps can be incorporated to lower the error on the testing dataset. Select
all that apply

☐ Try to obtain an additional set of features
☐ Try increasing the regularization parameter λ
☐ Get more training examples
☐ Try adding polynomial features

No, the answer is incorrect.
Score: 0

Accepted Answers:
Try to obtain an additional set of features
Get more training examples

7) Suppose you are training a regularized linear regression model. Check which of the following statements **1 point** are true? Select all that apply.

☐ The regularization parameter λ value is chosen so as to give the lowest training set error
☐ The regularization parameter λ value is chosen so as to give the lowest cross validation error
☐ The regularization parameter λ value is chosen so as to give the lowest test set error
☐ The performance of a learning algorithm on the training set will typically be better than its performance on the test set

No, the answer is incorrect.
Score: 0

Accepted Answers:
The regularization parameter λ value is chosen so as to give the lowest cross validation error
The performance of a learning algorithm on the training set will typically be better than its performance on the test set

8) ⬚   **1 point**

○ 1.03125
○ 2.03125
○ 3.03125
○ 4.03125

No, the answer is incorrect.
Score: 0

Accepted Answers:
2.03125

9)

1 point

○
○
○
○

No, the answer is incorrect.
Score: 0

Accepted Answers:

10) What is the cost now?

1 point

○ 0.4292
○ 1.4292
○ 2.4292
○ 3.4292

No, the answer is incorrect.
Score: 0

Accepted Answers:
0.4292

Previous Page   End

Unit 9 - Week 6

Assignment 6

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

Due on 2019-03-13, 23:59 IST.

1) Consider the following statements regarding Artificial Neural Network s(ANN) and Convolutional Neural Networks (CNN)   **1 point**

1. There are sparse connections between inputs and outputs between two consecutive layers in a CNN
2. Parameters are shared between output neurons in a CNN layer.
3. There are sparse connections between inputs and outputs between two consecutive layers of an ANN
4. For any two layers with the same number of neurons an ANN will have fewer parameters than a CNN

Which of the above statements are **TRUE**

- ○ 1 and 2
- ○ 1, 2 and 3
- ○ 1, 3 and 4
- ○ 2, 3 and 4

No, the answer is incorrect.
Score: 0

Accepted Answers:
1 and 2

2) What will be the size of the output of a convolutional layer with :   **1 point**

Input size = [227 x 227 x 3],
Filter Size = [11 x 11 x 3],
Stride = 4

- ○ [54 x 54]
- ○ [55 x 55]
- ○ [216 x 216]
- ○ [68 x 68]

No, the answer is incorrect.
Score: 0

Accepted Answers:
[55 x 55]

3) Pooling layers are used to accomplish which of the following?   **1 point**

- ☐ To progressively reduce the spatial size of the representation
- ☐ To reduce the amount of parameters and computation in network
- ☐ To select maximum value over pooling region always
- ☐ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
To progressively reduce the spatial size of the representation
To reduce the amount of parameters and computation in network

Answer questions 4-6 for the CNN architecture given below

The whole network is composed of CONV layers that perform 3x3 convolutions with stride 1 and padding is 'same'. POOL layers perform 2x2 max pooling with stride 2 (and no padding). Number of filters in the Conv layers and number of neurons in fully connected layers are shown in brackets

4)   **1 point**

The output size after pool1, pool2 are

- ○ [111x111x128], [56x56x64]
- ○ [112x112x128], [56x56x64]
- ○ [114x114x128], [58x58x64]
- ○ [111x111x128], [58x58x64]

No, the answer is incorrect.
Score: 0

5) Number of parameters till pool1 are *1 point*

- ○ 89186
- ○ 73570
- ○ 75648
- ○ 64898

No, the answer is incorrect.
Score: 0

6) Total number of parameters in the given network is *0 points*

- ○ 104097392
- ○ 206081344
- ○ 326789108
- ○ 207816190

No, the answer is incorrect.
Score: 0

7) Which of the following is true for most CNN architectures? *1 point*

- ☐ Size of input (height and width) decreases, while depth increases
- ☐ Multiple convolutional layers followed by pooling layers.
- ☐ Fully connected layers in the first few layers
- ☐ Fully connected layers in the last few layers
- ☐ Multiple pool layers followed by a convolutional layer

No, the answer is incorrect.
Score: 0

Consider the architecture shown below and answer Questions 8-10

8) *1 point*

The network shown is popularly known as

- ○ AlexNet
- ○ VGG
- ○ GoogLeNet
- ○ ResNet

No, the answer is incorrect.
Score: 0

9) What are the number of parameters and number of operations, for layer (3a) in above question? *1 point*

- ○ #parameters=163 K (approx.), #operations=128 M (approx.)
- ○ #parameters=159 K (approx.), #operations=128 M (approx.)
- ○ #parameters=128 M (approx.), #operations=159 K (approx.)
- ○ #parameters=128 K (approx.), #operations=159 M (approx.)

No, the answer is incorrect.
Score: 0

10) The importance of "reduce" in the table is that it *1 point*

- ○ Reduces no. of feature maps in the previous layer
- ○ Reduces no. of operations
- ○ Reduces no. of parameters
- ○ All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
All of the above

Unit 10 - Week 7

Assignment 7

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

Due on 2019-03-20, 23:59 IST.

Kindly download the paper from given link:

https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.

This paper describes Fully Convolutional Neural Networks (FCNNs) for semantic segmentation.
Read this paper and answer the following questions

1) The advantage of Fully Convolutional Networks over conventional CNNs   *1 point*

   a. Only fixed sized input data can be passed network.
   b. Any arbitrary sized input data can be passed through network.
   c. The size of the output of FCNNs depends on the input size.
   d. FCNN is a computationally cheaper way of achieving semantic segmentation compared to conventional CNNs

Which of the above statements are true?

○ a,b & d
○ a, c & d
○ b, c & d
○ a,b, c & d

No, the answer is incorrect.
Score: 0

Accepted Answers:
b, c & d

2) One issue in this specific FCN is that by propagating through several alternated convolutional and   *0 points*
pooling layers, the resolution of the output feature maps is down sampled. Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries. What methods are suggested by authors to gain original resolution?

☐ Bilinear Interpolation
☐ Deconvolution
☐ Backward convolution
☐ Forward convolution

No, the answer is incorrect.
Score: 0

Accepted Answers:
Bilinear Interpolation
Deconvolution
Forward convolution

3) ☐   *1 point*

○ 32x32x4096, 1x1x4096, 1x1x1000
○ 16x16x8192, 2x2x2048, 2x2x2048
○ 64x64x1024, 3x3x1000, 3x3x1000
○ 28x28x4096, 3x3x1024, 2x2x1024

No, the answer is incorrect.
Score: 0

Accepted Answers:
32x32x4096, 1x1x4096, 1x1x1000

4) Which of following statement/s is/are True?   *1 point*

☐ Decreasing subsampling within a net is a tradeoff: the filters see finer information, but have smaller receptive fields and take longer to compute.
☐ The shift-and-stitch trick is another kind of tradeoff: the output is denser without decreasing the receptive field sizes of the filters, but the filters are prohibited from accessing information at a finer scale than their original design.
☐ In this paper, network is trained with a per-pixel multinomial logistic loss.

☐ Network is validated with the standard metric of mean pixel intersection over union, with the mean taken over all classes, including background.

5) According to the paper, optimization parameter/s used in training the FCN-VGG16 is/are:　　*1 point*

☐ Stochastic Gradient descent used as optimization algorithms
☐ Learning rate is 10^{-4}
☐ For hyperparameter Grid Search has been used.
☐ Dropout has been used in the network.

6) In the context of deep learning, transfer learning depends on which of the following variables:　　*1 point*

☐ Similarity of the data to original data used for pre-training model weights
☐ Amount of data available
☐ Computational power
☐ Complexity of data

7) In this context of Transfer Learning which of the following statements are True?　　*1 point*

☐ Transfer learning enables feature extraction with pre-trained deep learning models.
☐ When target labels are scarce, the weights of pre-trained models are frozen (fix weights) so as to avoid overfitting.
☐ Fine-tuning of pre-trained model weights is generally preferred when target task labels are plentiful.
☐ Transfer learning works better when the tasks on which the networks are trained for are similar.

8) Choose the correct statement with regard to different approaches used for hyper-parameter optimisation.　　*1 point*

○ A. Grid search approach is usually very efficient, and each new guess is independent of previous guess.
○ B. Random search approach is usually more efficient than grid search, and each new guess is independent of previous guess.
○ C. In Bayesian optimisation, each new guess is independent of the previous guess.
○ Both A and B.

9) ▫　　*1 point*

○ High for model 1, low for model 2
○ Low for model 1, high for model 2

○ Low for both models
○ High for both models

No, the answer is incorrect.
Score: 0

Accepted Answers:
Low for model 1, high for model 2

Previous Page ◁ ▷ End

Unit 11 - Week 8

Course outline

How to access the portal

Matlab and Learning Modules

Pre-Requisite assignment

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

- ◯ Quiz : Assignment 8
- ◯ Activation Functions
- ◯ Learning Rate decay, Weight initialization
- ◯ Data Normalization
- ◯ Batch Norm
- ◯ Introduction to RNNs
- ◯ Example - Sequence Classification
- ◯ Training RNNs - Loss and BPTT
- ◯ Vanishing Gradients and TBPTT
- ◯ RNN Architectures
- ◯ LSTM
- ◯ Why LSTM Works
- ◯ Deep RNNs and Bi-RNNs
- ◯ Summary of RNNs
- ◯ Week 8 Feedback Form

Assignment 8

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

Due on 2019-03-27, 23:59 IST.

1) While training a vanilla RNN, the ML engineer finds that the weights keep growing with each epoch. Which of the following could be the possible reasons and solutions: **1 point**

- ☐ High learning rate. Try lowering learning rate.
- ☐ Not enough data. Get more training data.
- ☐ Bad architecture. Use LSTM
- ☐ Exploding gradients. Try clipping gradient.

No, the answer is incorrect.
Score: 0

Accepted Answers:
High learning rate. Try lowering learning rate.
Exploding gradients. Try clipping gradient.

2) Which of the following statements are true   **1 point**

- ◯ The forward pass of Vanilla RNNs is cheaper than that of GRUs and LSTMs
- ◯ Amongst GRUs, LSTMs and Vanilla RNNs, only LSTMs have a separate memory cell
- ◯ Usually, LSTMs can compute deeper sequences (without vanishing gradients) compared to GRUs
- ◯ All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
All of the above

3) What common property of AlexNet, LSTMs and ResNet helps in training?   **1 point**

- ◯ The number of their layers.
- ◯ The number of parameters
- ◯ They have multiple pathways for gradient backflow which helps in backprop
- ◯ All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
They have multiple pathways for gradient backflow which helps in backprop

A company is trying to automate case reports for MRI scans. The scans are videos of a beating heart at a particular cross section (slice). The automatic report is supposed to give a diagnosis amongst 5 different conditions (one of which is a "normal heart"). The videos are of 30 frames each. Each frame is a 227x227 grayscale image. Answer the following

4) Which is the most appropriate classification for the type of RNN to be used for this problem?   **1 point**

- ◯ One to One
- ◯ One to Many
- ◯ Many to One
- ◯ Many to Many

No, the answer is incorrect.
Score: 0

Accepted Answers:
Many to One

5) If we are to use the full, unprocessed video as the input to the RNN, which of the following are true?   **1 point**

- ☐ The unrolled RNN will have a depth of 30 in time.
- ☐ There are 30 sequential inputs to the RNN each of size 227x227
- ☐ There is only 1 non-sequential input to the RNN of size 227x227x30

Week 9

Week 10

Week 11

Week 12

Download Videos

TEXT
TRANSCRIPTS

Interaction session

☐   The RNN can be a deep RNN with 30 CNN like units

No, the answer is incorrect.
Score: 0

Accepted Answers:
The unrolled RNN will have a depth of 30 in time.
There are 30 sequential inputs to the RNN each of size 227x227
The RNN can be a deep RNN with 30 CNN like units

The ML engineer in the company decides that using the full, unprocessed video is too expensive for an RNN. So, she decides to encode the image by using a CNN architecture similar to Alexnet. The final, fully connected layer in her architecture has size 50. She uses this embedding (final layer) as the input to the RNN for each frame. She also uses a single hidden layer with 100 neurons. Answer the following questions.

6) What is the size of the matrix W_{hh}?   *1 point*

   ○ 5000x5000
   ○ 50x50
   ○ 100x100
   ○ 30x30

No, the answer is incorrect.
Score: 0

Accepted Answers:
100x100

7) What is the size of the matrix W_{xh}?   *1 point*

   ○ 50x100
   ○ 100x50
   ○ 100x100
   ○ 50x50

No, the answer is incorrect.
Score: 0

Accepted Answers:
100x50

8) What is the size of the bias matrix for the recurrent unit?   *1 point*

   ○ 50x1
   ○ 100x1
   ○ 30x1
   ○ 5000x1

No, the answer is incorrect.
Score: 0

Accepted Answers:
100x1

9) What is the size of the matrix W_{yh} if we are going to use softmax to classify? Ignore the bias unit here.   *1 point*

   ○ 500x1
   ○ 5x100
   ○ 100x5
   ○ 5x5

No, the answer is incorrect.
Score: 0

Accepted Answers:
5x100

10) What is the total number of parameters that we would have to train in case the RNN architecture used   *1 point*
was LSTM. Ignore all bias units as well as the output parameters

   ○ 15000
   ○ 15100
   ○ 45000
   ○ 60000

No, the answer is incorrect.
Score: 0

Accepted Answers:
60000

Courses » Machine Learning for Engineering and Science Applications

Unit 12 - Week 9

Assignment 9

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

Due on 2019-04-03, 23:59 IST.

1)   *1 point*

☐ ☐
☐ ☐
☐

☐ ☐

No, the answer is incorrect.
Score: 0

Accepted Answers:

☐

2) ☐   *1 point*

◯ ☐
◯ ☐
◯ ☐
◯ ☐

No, the answer is incorrect.
Score: 0

Accepted Answers:
☐

3) K-means is   *1 point*

◯ a. A probabilistic algorithm to identify clusters present in data
◯ b. A non-Probabilistic algorithm to identify clusters present in data
◯ None of the above
◯ Both a and b

No, the answer is incorrect.
Score: 0

Accepted Answers:
b. A non-Probabilistic algorithm to identify clusters present in data

4) Which of the following can act as possible termination conditions in K-Means?   *1 point*

- ○ Reaching a maximum number of iterations
- ○ Centroids do not change between successive iterations
- ○ The squared distance of each data point from its centroid summed over all training data points falls below a threshold
- ○ All of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
All of the above

5) Which of the following are true about Decision trees?   *1 point*

- ☐ Decision trees can be applied only for classification tasks
- ☐ Decision tree is a non-parametric method
- ☐ Decision tree can handle only categorical variables
- ☐ Decision trees tend to overfit data and are high variance classifiers

No, the answer is incorrect.
Score: 0

Accepted Answers:
Decision tree is a non-parametric method
Decision trees tend to overfit data and are high variance classifiers

6) You are given 1500 training data points to train a decision tree. The minimum number of observations in each child node should be 300 after a split at the parent node. The minimum number of data points falling into a leaf node should be at least 400. Given these conditions what is the maximum possible depth of the decision tree?   *1 point*

- ○ 1
- ○ 2
- ○ 3
- ○ 4

No, the answer is incorrect.
Score: 0

Accepted Answers:
3

7) Which of the following statements are True with regard to K-Nearest Neighbours?   *1 point*

- ☐ The decision boundary becomes smoother with decreasing value of K
- ☐ k-NN requires an explicit training step
- ☐ The K-Nearest Neighbor algorithm considers the entire training data for each test point classification
- ☐ Decreasing k increases variance

No, the answer is incorrect.
Score: 0

Accepted Answers:
The K-Nearest Neighbor algorithm considers the entire training data for each test point classification
Decreasing k increases variance

8) Download the Old Faithful data from the following link   *1 point*

http://www.stat.cmu.edu/~larry/all-of-statistics/=data/faithful.dat

Assume, you want to cluster the given data set into 2 clusters, using K-Means clustering algorithm. What will be the cluster centroids, on convergence?

- ○ C1: (2.09, 54.75) , C2: (4.29,80.28)
- ○ C1: (1.88, 55), C2: (4.15, 88)
- ○ C1: (2.18, 55), C2: (4.80, 81)
- ○ None of the above

No, the answer is incorrect.

Accepted Answers:
C1: (2.09, 54.75) , C2: (4.29,80.28)

9) Which of the following is true with respect to bagging?   *1 point*

☐ Bagging involves sampling unbiasedly from the data for the purpose of ensemble learning.
☐ Bagging involves using weights on particular data points while sampling.
☐ Bagging typically reduces variance in the trained model.
☐ Bagging typically reduces bias in the trained model.

No, the answer is incorrect.
Score: 0

Accepted Answers:
Bagging involves sampling unbiasedly from the data for the purpose of ensemble learning.
Bagging typically reduces variance in the trained model.

10) Which of the following is true with respect to boosting?   *1 point*

☐ Boosting can lead to overfitting the data
☐ Boosting only reduces variance in the trained model
☐ Boosting tries to bias the overall model by weighting in the favor of good performers
☐ Boosting involves unbiased sampling of the data

No, the answer is incorrect.
Score: 0

Accepted Answers:
Boosting can lead to overfitting the data
Boosting tries to bias the overall model by weighting in the favor of good performers

Previous Page   End

Unit 13 - Week 10

Assignment 10

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

Due on 2019-04-10, 23:59 IST.

1) Which of the following is assumed when classifying using the Naive Bayes method?   *1 point*

- ○ The features have strong correlation with each other
- ○ The features are all normalized before classification
- ○ The features are independent/conditionally independent of each other
- ○ The data is low dimensional in nature

No, the answer is incorrect.
Score: 0

Accepted Answers:
The features are independent/conditionally independent of each other

2) The primary role of Principal component analysis is   *1 point*

- ○ Classification
- ○ Regression
- ○ Clustering
- ○ Dimensionality Reduction

No, the answer is incorrect.
Score: 0

Accepted Answers:
Dimensionality Reduction

3) Which direction does PCA use to perform dimensionality reduction?   *1 point*

- ○ Direction of maximum variance
- ○ Direction of minimum variance
- ○ Direction of maximum mean
- ○ Direction of minimum mean

No, the answer is incorrect.
Score: 0

Accepted Answers:
Direction of maximum variance

4) Which of the following is/are true about PCA?   *1 point*

- ☐ The principal components are orthogonal to each other
- ☐ There can be only one principal component in PCA
- ☐ There are always lesser principal components than initial dimensions
- ☐ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
The principal components are orthogonal to each other
There are always lesser principal components than initial dimensions

5)   *1 point*

- ○ [   ]
- ○ [   ]
- ○ [   ]
- ○ [   ]

No, the answer is incorrect.
Score: 0

Accepted Answers:
[   ]

The outcome of a series of N tosses is given by the vector X which contains either 1 or 0 in each element. Assuming that the probability of a heads is μ and that each toss is independent, the probability $P(X|\mu)$ is given by

By above data answer the following questions 6 &7

6) *1 point*

○

○

○

○

No, the answer is incorrect.
Score: 0

Accepted Answers:

7) *1 point*

○

○

○

○

No, the answer is incorrect.
Score: 0

Accepted Answers:

8) Which of the following is true for the Gaussian distribution? *1 point*

☐ The mean of the distribution is always zero
☐ In higher dimensions, covariance is represented by a matrix
☐ The Gaussian is always symmetric about the mean
☐ The mean of the Gaussian is the same as its mode

No, the answer is incorrect.
Score: 0

Accepted Answers:
In higher dimensions, covariance is represented by a matrix
The Gaussian is always symmetric about the mean
The mean of the Gaussian is the same as its mode

9) During linear regression, the maximum likelihood estimate of the parameters would be *1 point*

○ Greater than the least squares parameters
○ Lesser than the least squares parameters
○ Same as the least squares parameters
○ Have no connection with the least squares parameters

No, the answer is incorrect.
Score: 0

Accepted Answers:
Same as the least squares parameters

10) Suppose you wanted to model the probability distribution of a set of people liking a particular dish. Which form of *1 point*
distribution would be most apt? (Hint: They can either like the dish or they can't.)

○ Normal Distribution
○ Bernoulli Distribution
○ Exponential Distribution
○ Log-Normal Distribution

No, the answer is incorrect.
Score: 0

Accepted Answers:
Bernoulli Distribution

Courses » Machine Learning for Engineering and Science Applications

Announcements   Course   Ask a Question   Progress   FAQ

Unit 14 - Week 11

Course outline

How to access the portal

Matlab and Learning Modules

Pre-Requisite assignment

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Week 11

- ○ Quiz : Assignment 11

- ○ MLE, MAP and Bayesian Regression

- ○ Introduction to Generative model

- ○ Generative Adversarial Networks (GAN)

- ○ Variational Auto-encoders (VAE)

- ○ Applications: Cardiac MRI - Segmentation & Diagnosis

- ○ Applications: Cardiac MRI Analysis - Tensorflow code walkthrough

- ○ Machine Learning for Engineering and Science Applications : Week 11 Feedback Form

Assignment 11

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

Due on 2019-04-17, 23:59 IST.

1)   *1 point*

○

○

○

○

No, the answer is incorrect.
Score: 0

Accepted Answers:

2)   *1 point*

○

○

○

○

No, the answer is incorrect.
Score: 0

Accepted Answers:

3) For any particular problem, maximizing the likelihood function always leads to   *1 point*

○ High Bias
○ Over-fitting
○ Simple models

○ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
Over-fitting

4) *1 point*

☐

☐

☐

☐

No, the answer is incorrect.
Score: 0

Accepted Answers:

5) What does an auto-encoder learn about the data? *1 point*

○ High dimensional representation of the data
○ Low dimensional representation of the data
○ Average dimensional representation of the data
○ No representation of the data is learned

No, the answer is incorrect.
Score: 0

Accepted Answers:
Low dimensional representation of the data

6) Auto-encoders are able to compress the input data in its hidden representation if: *1 point*

○ If the input features are correlated
○ If the input features are not correlated
○ If the input features are independent
○ If the input features are unrelated

No, the answer is incorrect.
Score: 0

Accepted Answers:
If the input features are correlated

7) Generative Adversarial models are: *1 point*

○ A. Generative models
○ B. Discriminative models
○ Both A and B
○ None of the above.

No, the answer is incorrect.
Score: 0

Accepted Answers:
A. Generative models

8) The main objective of Bayesian linear regression is *1 point*

○ A. To find a single best value of model parameters.
○ B. To determine the posterior distribution for the model parameters
○ Both A and B
○ None of the above.

No, the answer is incorrect.
Score: 0

Accepted Answers:
B. To determine the posterior distribution for the model parameters

9)   *1 point*

☐ [ ]
☐ [ ]
☐ [ ]
☐ [ ]

No, the answer is incorrect.
Score: 0

Accepted Answers:

[ ]

[ ]

10) Which of the following are true about variational auto-encoders?   *1 point*

○ They are a form of denoising autoencoders
○ They require a random number as input during training
○ It's called variational because the output varies constantly
○ KL divergence is one of the terms in the loss function

No, the answer is incorrect.
Score: 0

Accepted Answers:
KL divergence is one of the terms in the loss function

Previous Page   End

Courses » Machine Learning for Engineering and Science Applications

Announcements   Course   Ask a Question   Progress   FAQ

Unit 15 - Week 12

Assignment 12

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

Due on 2019-04-24, 23:59 IST.

This week's assignments are all based on reading and understanding some application-based papers which we discussed during the lectures.

Download https://arxiv.org/pdf/1711.10561.pdf and answer the following questions (1-5)

1) In the paper what does MSE_u refer to?   **1 point**

- ○ The mean squared error for the boundary data
- ○ The mean squared error of the differential equation's residual
- ○ The mean squared error of the initial data
- ○ The mean squared error of the boundary and initial data combined

No, the answer is incorrect.
Score: 0

Accepted Answers:
The mean squared error of the boundary and initial data combined

2) In the paper what does MSE_f refer to?   **1 point**

- ○ The mean squared error for the boundary data
- ○ The mean squared error of the differential equation's residual
- ○ The mean squared error of the initial data
- ○ The mean squared error of the boundary and initial data combined

No, the answer is incorrect.
Score: 0

Accepted Answers:
The mean squared error of the differential equation's residual

3) We wish to solve Laplace's equation $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}$=0 using the PINN approach. The appropriate neural network structure would be   **1 point**

- ○ A CNN, as it is in 3D
- ○ An ANN with 3 inputs and 3 outputs and one hidden layer
- ○ An ANN with 3 inputs and 1 output and one hidden layer
- ○ An ANN with 3 inputs and 1 output and as many hidden layers as we want

No, the answer is incorrect.
Score: 0

Accepted Answers:
An ANN with 3 inputs and 1 output and as many hidden layers as we want

4) Which of the following is true of PINN?   **1 point**

- ☐ The boundary conditions are satisfied exactly
- ☐ The boundary conditions are satisfied only approximately
- ☐ The PDE/ODE is converted to an optimization problem
- ☐ It uses automatic differentiation in order to compute the derivatives in the ODE/PDE

No, the answer is incorrect.
Score: 0

Accepted Answers:
The boundary conditions are satisfied only approximately
The PDE/ODE is converted to an optimization problem
It uses automatic differentiation in order to compute the derivatives in the ODE/PDE

5) Suppose our differential equation had the term $\frac{\partial u}{\partial x \partial t}$ as well. Within the code snippet for "def f(t,x)"   **1 point**

Summary and road ahead

Quiz : Assignment 12

Week 12 Feedback Form

Download Videos

TEXT TRANSCRIPTS

Interaction session

given in the paper, which of the following would compute the term $\dfrac{\partial u}{\partial x \partial t}$ for the neural network?

- [ ] u_xt = tf.gradient(u_x, t)[0]
- [ ] u_xt = tf.gradient(u_t, x)[0]
- [ ] u_xt = tf.gradient(u, x, t)[0]
- [ ] u_xt = tf.gradient(u_xt)[0]

No, the answer is incorrect.
Score: 0

Accepted Answers:
u_xt = tf.gradient(u_x, t)[0]
u_xt = tf.gradient(u_t, x)[0]

Download https://www.autodeskresearch.com/sites/default/files/ADSK-KDD2016.pdf and answer the following questions (6-10)

6) The learning approach utilized in this paper is   *1 point*

- ( ) Supervised learning
- ( ) Semi-supervised learning
- ( ) Unsupervised learning
- ( ) Reinforcement learning

No, the answer is incorrect.
Score: 0

Accepted Answers:
Supervised learning

7) The authors tried which of the following experiments in the paper (Mark all that are true)   *1 point*

- [ ] Single encoder with separate decoders for each velocity component
- [ ] Separate decoders for each velocity component
- [ ] The input layer being defined by a binary pixel value (0 inside the body and 1 outside)
- [ ] The input layer being defined by a signed distance function

No, the answer is incorrect.
Score: 0

Accepted Answers:
Single encoder with separate decoders for each velocity component
Separate decoders for each velocity component
The input layer being defined by a binary pixel value (0 inside the body and 1 outside)
The input layer being defined by a signed distance function

8) Why do the authors perform a patch-wise linear regression computation?   *1 point*

- ( ) For hyperparameter optimization
- ( ) For improving their predictions via boosting
- ( ) In order to see if their architecture is better than a simple locally linear interpolation
- ( ) In order to reduce their computational expense on GPUs

No, the answer is incorrect.
Score: 0

Accepted Answers:
In order to see if their architecture is better than a simple locally linear interpolation

9) Which of the following is true of the 2D geometries in the paper?   *1 point*

- [ ] The stride in the first layer is 16x8
- [ ] There is a single loss function for both the velocity components
- [ ] The strides are the same size as the horizontal and vertical filter sizes
- [ ] There is a separate loss function for each velocity component

No, the answer is incorrect.
Score: 0

Accepted Answers:
The stride in the first layer is 16x8
There is a single loss function for both the velocity components
The strides are the same size as the horizontal and vertical filter sizes

10) Which of the following is true of the CNN based solution used in this paper?   *1 point*

- [ ] Could be used for initial design of shapes of cars as it is faster
- [ ] It is more accurate than traditional solutions
- [ ] It requires a large database of existent CFD solutions
- [ ] All of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:
Could be used for initial design of shapes of cars as it is faster
It requires a large database of existent CFD solutions

Previous Page

End