Machine Learning (CS1741) Quiz-I

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

Points: 5/5

1. Consider the following statements and which of the following error function suits best for it?
(1/1 Point)

$Given\ (x_1, t_1), (x_2, t_2) \ldots . (x_N, t_N), best\ fitting\ data\ to\ y(X, W) = W^T X + w_0\ by\ least\ square$

- $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- $\sum_{i=1}^{N} |t_i - y(x_i, W)|$
- $\sum_{i=1}^{N} [t_i - y(x_i, W)]^2$ ✓
- $None\ of\ these$

2. The polynomial curve fitting often suffers from over-fitting problem.
(1/1 Point)

- True ✓
- False

3. To find the minimum or the maximum of a function, we set the gradient to zero because:
(1/1 Point)

- (●) The value of the gradient at extreme of a function is always zero ✓
- ( ) Depends on the type of problem
- ( ) Both are correct
- ( ) None of these

4. Supervised learning differs from unsupervised learning in that supervised learning requires _____.
(1/1 Point)

labelled data while training an algorithm

Correct answers: Labeled/Trainig data

5. Which one of the followings is true?
(1/1 Point)

- ( ) Supervised learning is good for clustering problems
- (●) Unsupervised learning does not require target values ✓
- ( ) Both are correct
- ( ) None of the above

Machine Learning (CS1741) Quiz-I

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

Points: 5/5

1. Output of a linear regression model is continuous, whereas output of classification is discrete.
(1/1 Point)

- (●) True ✓
- ( ) False

2. Suppose you are training a machine learning model for binary classification. What type of error function which is suitable to solve the problem.
(1/1 Point)

$Given\ (x_1, t_1), (x_2, t_2) \ldots . (x_N, t_N), best\ fitting\ data\ to\ y(X, W) = W^T X + w_0\ by\ least\ squares.$

- ( ) $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- ( ) $\sum_{i=1}^{N} |t_i - y(x_i, W)|$
- (●) $\sum_{i=1}^{N} [t_i - y(x_i, W)]^2$ ✓
- ( ) $None\ of\ these$

3. If the training examples are linearly separable, how many decision boundaries can separate positive from negative data points?
(1/1 Point)

- One
- Two
- Infinite ✓
- None of these

4. The perceptron learning suffers from classifying non-linear data sets and it can't handle _____.
(1/1 Point)

- Overfitting
- Underfitting
- Noise ✓
- None of these

5. Which of the following sentence is FALSE regarding regression?
(1/1 Point)

- It relates inputs to outputs
- It is used for prediction
- It may be used for interpretation
- It discovers causal relationships ✓

Machine Learning (CS1741) Quiz-I

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

Points: 5/5

1. If you are using all features of my dataset and you achieve 100% accuracy on my training set using a machine learning model, but ~70% on test data set, what should you look out for?
(1/1 Point)

- Underfitting
- Nothing, the model is perfect
- Overfitting ✓
- None of the above

2. Which of the following is true about Unsupervised learning?
(1/1 Point)

- Categorizes training vectors by identifying similarities between them ✓
- Can use the same error functions as supervised learning
- It is collaborative learning methods which often applied between classes
- None of the above

3. The k-means algorithm is _____ learning algorithm.
(1/1 Point)

Unsupervised

Correct answers: an unsupervised

4. Suppose you are training a machine learning model for binary classification. What type of error function which is suitable to solve the problem.
(1/1 Point)

Given $(x_1, t_1), (x_2, t_2) \ldots . (x_N, t_N)$, *best fitting data to* $y(X, W) = W^T X + w_0$ *by least squares.*

- $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- $\sum_{i=1}^{N} |t_i - y(x_i, W)|$
- $\sum_{i=1}^{N} [t_i - y(x_i, W)]^2$ ✓
- *None of these*

5. Classification concerns finding decision boundaries that can be used to separate out different classes.
(1/1 Point)

- True ✓
- False

Machine Learning (CS1741) Quiz-II

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

Points: 3/5

1

Which of the following is true about multi-layer perceptron network?
(0/1 Point)

- Usually, the weights are initially set to small random values. ✓
- A hard-limiting activation function is often used.
- The weights can only be updated after all the training vectors have been presented.
- Multiple layers of neurons allow for less complex decision boundaries than a single layer.

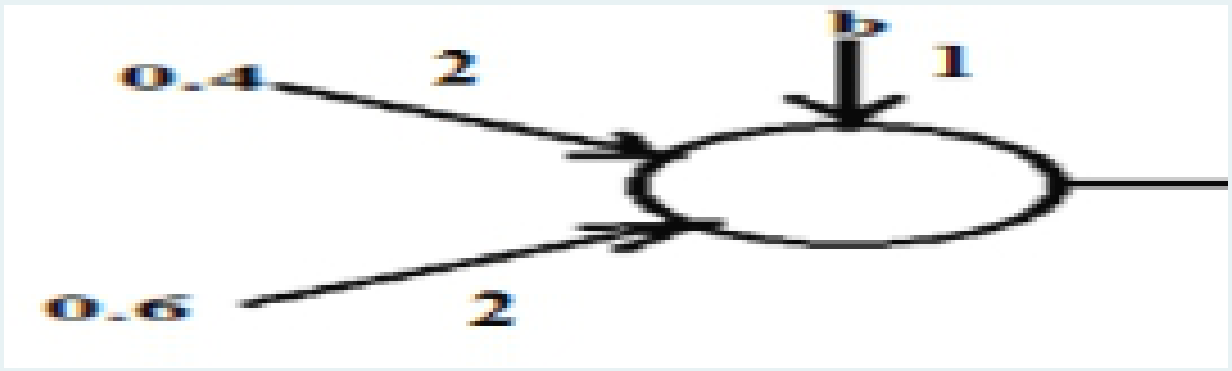
2

Suppose we train a hard-margin linear SVM on n > 100 data points, yielding a hyperplane with exactly 2 support vectors. If we add one more data point and retrain the classifier, what is the maximum possible number of support vectors for the new hyperplane (assuming the n+1 points are linearly separable)?
(1/1 Point)

- 2
- 3
- n
- n+1 ✓

3

What will be the minimum value of Bias (b), if binary step function is used as activation function with threshold 2.3 and target is 1 for the following network ?
(0/1 Point)



- 0.4
- 0.3 ✓
- 0.5
- None of these

4

Which of the following is true about Support Vector Machines (SVMs)?
(1/1 Point)

- Support vectors are used for computing hyperplanes. ✓
- It is a method for minimizing the margin to hyperplanes.

- Nonlinear problems are handled with mapping inputs to lower-dimensional space.
- None of the above

5

The kernel trick
(1/1 Point)

- can be applied to every classification algorithm.
- changes ridge regression so we solve a d x d linear system instead of an n x n system, given n sample points with d features.
- is commonly used for dimensionality reduction.
- exploits the fact that in many learning algorithms, the weights can be written as a linear combination of input points. ✓

Machine Learning (CS1741) Quiz-II

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

...

Points: 4/5

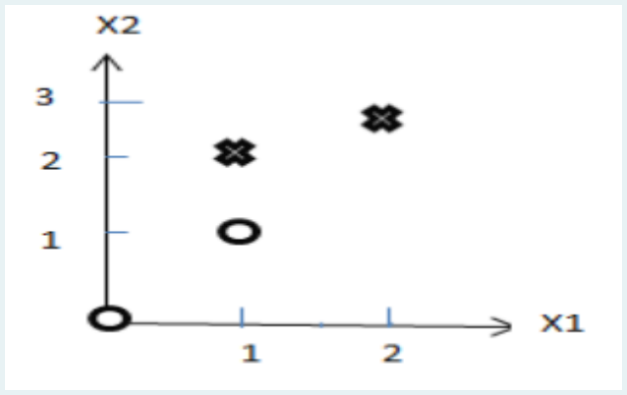
1

A perceptron with 4 inputs has the weight vector w = [1 2 3 4]T. The activation function is linear and given by f(yin) = 2yin. If the input vector is X = [5 6 7 8] T, then find the output of the neuron.
(1/1 Points)

- 70
- 128
- 140 ✓
- 64

2

Consider the following two-class data set as shown in Figure. Just by visual inspection, find the decision boundary learnt by SVM. [X - Class 1 and O - Class 0].
(1/1 Points)



- $X_1 = 1.5$
- $X_2 = 1.5$ ✓
- $X_1 + X_2 = 1.5$
- *None of these*

✕

3

Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0, 1].
(0/1 Points)

- Sigmoid ✓
- Mode
- Square
- All of the above

4

Which of the following is true about classification ?
(1/1 Points)

- It concerns finding decision boundaries that can be used to separate out different classes.
- Non-linear decision boundaries can solve more complex problems than linear boundaries (straight lines).
- A test set is more relevant for testing generalization than the training set.
- All of the above. ✓

5

What do you mean by a hard margin in SVM?
(1/1 Points)

- The SVM allows very low error in classification. ✓
- (B)The SVM allows high amount of error in classification.
- The SVM allows moderate error in classification.
- None of the above

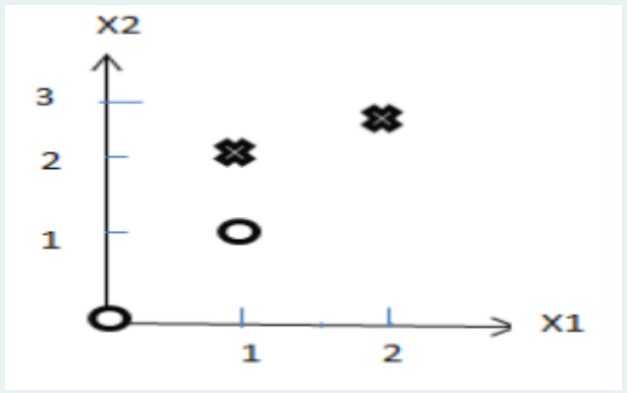
Machine Learning (CS1741) Quiz-II

Total Marks: 5 Each question carries equal marks (1).
Time: 10 Mints

...

Points: 4/5

1



Consider the following two-class data set as shown in Figure. Just by visual inspection, find the decision boundary learnt by SVM. [X - Class 1 and O - Class 0].
(1/1 Points)

- $X_1 = 1.5$
- $X_2 = 1.5$ ✓
- $X_1 + X_2 = 1.5$
- None of these

2

Consider a following model for logistic regression:
(1/1 Points)

$P(y = 1|x, w) = g(w_0 + w_1 x)$ *where* $g(z)$ *is the* log *istic function. What would be the value of*

- $(0, \infty)$
- $(-\infty, 0)$
- $(0, 1)$ ✓
- $(-\infty, \infty)$

✗

3

Find a suitable kernel K that counts the number of position-wise matches between two DNA sequences. For instance, K (ACTGG, ATCG) = 2, and K(AACTCG, ACCTGGA) = 4.
(0/1 Points)

- $K(x, z) = \phi(x) . \phi(z)$
- $K(x, z) = \phi(x) \ \phi(z)^T$
- Both A and B ✓
- *None of the these*

4

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 are
(1/1 Points)

- 50 ✓
- *Less than* 50
- *More than* 50
- It is an arbitrary value

5

A feed-forward neural network is said to be fully connected when
(1/1 Points)

- All nodes are connected to each other.
- All nodes at the same layer are connected to each other.
- All nodes at one layer are connected to all nodes in the next higher layer. ✓
- All hidden layer nodes are connected to all output layer nodes.

Machine Learning (CS1741) (EL VII) Sessional I

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

Points: 15/15

1. Which of the following error function suits best for least squares regression? (1/1 Point)

$Given\ (x_1, t_1),\ (x_2, t_2) \ldots . (x_N, t_N),\ best\ fitting\ data\ to\ y(X, W) = W^T X + w_0\ by\ leastsquare$

- [x] $\frac{1}{2} \sum_{i=1}^{N} [t_i - y(x_i, W)]^2$
- [ ] $None\ of\ these$
- [ ] $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- [ ] $\sum_{i=1}^{N} |t_i - y(x_i, W)|$

2. Which of the following is true about Unsupervised learning? (1/1 Point)

- [ ] It is collaborative learning methods which often applied between classes
- [ ] Can use the same error functions as supervised learning
- [ ] None of the above
- [x] Categorizes training vectors by identifying similarities between them

3. The k-means algorithm is _____ learning algorithm.
(1/1 Point)

- a supervised
- None of the above
- a semi-supervised
- an unsupervised ✓

4. Suppose you have a two-class data set in 2-D space, generated as follows. Obtain the feature transform function) so that a linear SVM can classify the data points correctly.
(2/2 Points)

Positive samples and negative taken from points on the curve $x_1^2 + x_2^2 = 5$, *and* $x_1^2 + x_2^2 = 1$

- $f(x) = x_1 x_2$
- *None of these*
- $f(x) = x_1^2 x_2^2$
- $f(x) = x_1^2 + x_2^2$ ✓

5. The perceptron learning suffers from classifying non-linear data sets and it can't handle _____.
(1/1 Point)

- Underfitting
- Overfitting
- None of these
- Noise ✓

6. If you are using all features of my dataset and you achieve 100% accuracy on my training set using a Multi-layer Neural Network model, but ~70% on validation set, what should you look out for?
(1/1 Point)

- *Underfitting*
- *None of the above*
- *Overfitting* ✓
- *Nothing, the model is perfect*

7. Adding more basis functions in a linear model... (pick the most probably option)
(1/1 Point)

- Doesn't affect bias and variance
- Decreases estimation bias
- Decreases variance
- Decreases model bias ✓

8. The polynomial curve fitting often suffers from over-fitting problem.
(1/1 Point)

- True ✓
- False

9. Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)?
S1: In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.
S2: In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.
S3: In GD, you either use the entire data or a subset of training data to update a parameter in each iteration.
(2/2 Points)

- *Only S3*
- *S1 and S2*
- *Only S1* ✓
- *Only S2*

10. Which of the following is true for classification
(1/1 Point)

- It concerns finding decision boundaries that can be used to separate out different classes
- Non-linear decision boundaries can solve more complex problems than linear boundaries (straight lines)
- A test set is more relevant for testing generalization than the training set
- All of the above. ✓

11. If the training examples are linearly separable, how many decision boundaries can separate positive from negative data points?
(1/1 Point)

- One
- Infinite ✓
- Two
- None of these

12. A perceptron with 4 inputs has the weight vector W. The activation function is linear and given by f(yin) = 2yin. If the input vector is X, then find the output of the neuron given the followings.
(2/2 Points)

$W = [1\ 2\ 3\ 4]^T$ *and* $X = [5\ 6\ 7\ 8]^T$

- 70
- 64
- 128
- 140 ✓

Machine Learning (CS1741) (EL VII) Sessional I

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

...

Points: 14/15

1. Adding more basis functions in a linear model... (pick the most probably option) (1/1 Points)

   - (●) Decreases model bias ✓
   - ( ) Decreases estimation bias
   - ( ) Decreases variance
   - ( ) Doesn't affect bias and variance

2. The polynomial curve fitting often suffers from under-fitting problem. (1/1 Points)

   - ( ) True
   - (●) False ✓

3. Suppose, you applied a Linear Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is correct in such a case. (2/2 Points)

- ○ Training accuracy increases
- ○ Testing accuracy increases or remains the same
- ◉ Both are correct ✓
- ○ None of these

4. 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 are (2/2 Points)

- ◉ 50 ✓
- ○ *Less than* 50
- ○ *More than* 50
- ○ *It is an arbitrary value*

5. Which of the following statements about regularization is not correct? (0/1 Points)

- ◉ Using too large a value of lambda can cause your hypothesis to underfit the data.
- ○ Using too large a value of lambda can cause your hypothesis to overfit the data.
- ○ Using a very large value of lambda cannot hurt the performance of your hypothesis.
- ○ None of the above ✓

6. The perceptron learning suffers from classifying non-linear data sets and it can't handle _____.
(1/1 Points)

- Overfitting
- Underfitting
- Noise ✓
- Outliers

7. Which of the following is a good test dataset characteristic?
(1/1 Points)

- Large enough to yield meaningful results
- Is representative of the dataset as a whole
- Both are correct ✓
- None of the above

8. Supervised learning differs from unsupervised learning in that supervised learning requires _____.
(1/1 Points)

- Labelled data
- Training data
- Both are correct ✓
- None of the above

9. Suppose you are training a multilayer neural network for binary classification. What type of error function which is suitable to solve the problem. (1/1 Points)

- $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- $\sum_{i=1}^{N} |t_i - y(x_i, W)|$
- $\frac{1}{2} \sum_{i=1}^{N} [t_i - y(x_i, W)]^2$ ✓
- *None of these*

10. Which of the following sentence is FALSE regarding regression? (1/1 Points)

- It relates inputs to outputs.
- It is used for prediction.
- It may be used for interpretation.
- It discovers causal relationships. ✓

11. (2/2 Points)

The log *isticfunction is given by* $f(x) = \frac{1}{1+e^{-x}}$ *Which of the following is correct*?

- $f'(x) = f(x)(1 - f(x))$ ✓
- $f'(x) = f(x)(1 + f(x))$
- $f'(x) = f(x)(1 - f(x))^2$
- *None of these*

12. To find the minimum or the maximum of a function, we set the gradient to zero because:
(1/1 Points)

- (●) The value of the gradient at extrema of a function is always zero ✓
- ( ) Depends on the type of problem
- ( ) Both are correct
- ( ) None of these

Machine Learning (CS1741) (EL VII) Sessional I

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

...

Points: 14/15

1. The k-means algorithm is _____ .
(1/1 Points)

- a supervised learning algorithm.
- an unsupervised learning algorithm. ✓
- a semi-supervised learning algorithm.
- a weakly supervised learning algorithm.

2. Assume a simple MLP model with 3 neurons and inputs= 1,2,3. The weights to the input neurons are -4, -5 and 6 respectively. Assume the activation function, f(x) = 3x. What will be the output?
(2/2 Points)

- 5
- 4 ✓

- 6
- 10

3. Which of the following is true about Naive Bayes?
(1/1 Points)

- Assumes that all the features in a dataset are equally important
- Assumes that all the features in a dataset are independent
- Both are correct ✓
- None of the above

4. Which of the following is true about k-means algorithm?
(0/1 Points)

- Automatically finds the number of clusters
- Each cluster centre is moved to the mean of data points assigned to it for each iteration
- A too small number of clusters may lead to overfitting
- None of the above ✓

5. Which of the following is true about Unsupervised learning?
(1/1 Points)

- Categorizes training vectors by identifying similarities between them ✓
- Can use the same error functions as supervised learning
- It is collaborative learning methods which often applied between classes
- None of the above

6. Which of the following error function suits best for least squares regression?
(2/2 Points)

$Given\ (x_1, t_1),\ (x_2, t_2) \ldots . (x_N, t_N),\ best\ fitting\ data\ to\ y(X, W) = W^T X + w_0\ by\ leastsquares.$

- $\sum_{i=1}^{N} [t_i - y(x_i, W)]$
- $\sum_{i=1}^{N} |t_i - y(x_i, W)|$
- $\frac{1}{2} \sum_{i=1}^{N} [t_i - y(x_i, W)]^2$ ✓
- $None\ of\ these$

7. In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden output layer and input hidden layer?
(2/2 Points)

- $[8 \times 5]\ and\ [5 \times 1]$ ✓
- $[5 \times 8]\ and\ [5 \times 1]$
- $[8 \times 5]\ and\ [1 \times 5]$
- $[5 \times 8]\ and\ [1 \times 5]$

8. The polynomial curve fitting often suffers from_____ problem.
(1/1 Points)

- Underfitting
- Overfitting ✓
- Both are correct
- None of the above

9. Which one of the followings is true?
(1/1 Points)

- Supervised learning is good for clustering problems
- Unsupervised learning does not require target values ✓
- Both are correct
- None of the above

10. If the training examples are linearly separable, how many decision boundaries can separate positive from negative data points?
(1/1 Points)

- *One*
- *Two*
- *Infinite* ✓
- *None of these*

11. What is Machine learning?
(1/1 Points)

- The autonomous acquisition of knowledge using computer programs ✓
- The autonomous acquisition of knowledge using manual programs
- The selective acquisition of knowledge using computer programs
- The selective acquisition of knowledge using manual programs

12. Output of a linear regression model is continuous, whereas output of classification is discrete.
(1/1 Points)

- (●) True ✓
- ( ) False

Machine Learning (CS1741) (EL VII) Sessional II

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

Points: 11/15

1. Which one of the followings is correct about Neural networks
(0/1 Point)

- Neural networks optimize a convex cost function
- Neural networks can be used for regression as well as classification ✓
- Neural networks always output values between 0 and 1
- None of the above

2. Suppose you have trained a logistic regression classifier and it outputs a new example x with a prediction ho(x) = 0.2. This means
(2/2 Points)

- Estimation for P (y=1 | x)
- Estimation for P (y=0 | x) ✓
- Estimation for P (y=1 | x)

- Estimation for P(y=0 | x)

3. Which of the following methods do we use to best fit the data in Logistic Regression?
(1/1 Point)

- Least Square Error
- Maximum Likelihood ✓
- Both A and B
- None of these

4. The kernel trick
(1/1 Point)

- is commonly used for dimensionality reduction.
- exploits the fact that in many learning algorithms, the weights can be written as a linear combination of input points. ✓
- can be applied to every classification algorithm
- changes ridge regression so we solve a d x d linear system instead of an n x n system, given n sample points with d features.

5. In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden output layer and input hidden layer?
(0/1 Point)

- $[8 \times 5]$ *and* $[5 \times 1]$ ✓
- $[5 \times 8]$ *and* $[5 \times 1]$
- $[8 \times 5]$ *and* $[1 \times 5]$
- $[5 \times 8]$ *and* $[1 \times 5]$

6. In neural networks, nonlinear activation functions such as sigmoid, tanh, and ReLU
(0/1 Point)

- speed up the gradient calculation in back-propagation, as compared to linear units.
- are applied only to the output units.
- help to learn nonlinear decision boundaries. ✓
- always output values between 0 and 1.

7. 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 are
(2/2 Points)

- 50 ✓
- Less than 50
- More than 50
- It is an arbitrary value

8. Which of the following are true of convolutional neural networks (CNNs) for image analysis?
(1/1 Point)

- Pooling layers reduce the spatial resolution of the image ✓
- They have more parameters than fully connected networks with the same number of layers and the same numbers of neurons in each layer
- A CNN can be trained for unsupervised learning tasks, whereas an ordinary neural net cannot.
- None of these.

9. Which of the following is true about multilayer perceptron network?
(1/1 Point)

- (●) Usually, the weights are initially set to small random values ✓
- ( ) A hard-limiting activation function is often used
- ( ) Multiple layers of neurons allow for less complex decision boundaries than a single layer
- ( ) None of the above

10. Find a suitable kernel K that counts the number of position-wise matches between two DNA sequences. For instance, K (ACTGG, ATCG) = 2, and K(AACTCG, ACCTGGA) = 4.
(0/1 Point)

- ( ) $K\ (x,\ z)\ =\ \phi\ (x)\ .\ \phi(x)$
- (●) $K\ (x,\ z)\ =\ \phi\ (x)\ .\ \phi(x)^T$
- ( ) *Both A and B are correct* ✓
- ( ) *None of the above*

11. The perceptron learning suffers from classifying non-linear data sets and it can't handle _____.
(1/1 Point)

- ( ) Overfitting
- ( ) Underfitting
- (●) Noise ✓
- ( ) None of these

12. Which of the following is correct for the given logistic function?
(2/2 Points)

$f(x) = \frac{1}{1+e^{-x}}$

- $\sigma'(x) = \sigma(x)(1-\sigma(x))$ ✓
- $\sigma'(x) = \sigma(x)(1+\sigma(x))$
- $\sigma'(x) = \sigma(x)(1-\sigma(x))^2$
- *None of the above*

Machine Learning (CS1741) (EL VII) Sessional II

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

...

Points: 15/15

1

Hard margin in SVM allows very low error in classification.
(1/1 Points)

- (●) True ✓
- ( ) False

2

A perceptron with 4 inputs has the weight vector W. The activation function is linear and given by f(yin) = 2yin. If the input vector is X, then find the output of the neuron given the followings.
(2/2 Points)

$W = [1\ 2\ 3\ 4]^T\ and\ X = [5\ 6\ 7\ 8]^T$

- ( ) 70

- 128
- 140 ✓
- 64

3

Suppose you are applying 5x5 convolutional mask on 32x32 input image (neurons) from MNIST data set. Compute the size of the first convolved layer/hidden on valid convolution without zero padding. The size of the stride is given as 1.
(2/2 Points)

- 30×30
- 28×28 ✓
- 32×32
- 26×26

4

If the training examples are linearly separable, how many decision boundaries can separate positive from negative data points in Support Vector Machines (SVM)?
(1/1 Points)

- One
- Two
- Infinite ✓
- None of these

5

Consider a following model for logistic regression: P (y =1|x, w)= g(w0 + w1x) where g(z) is the logistic function. What would be the value of P ?
(1/1 Points)

- $(0,\ \infty)$
- $(-\infty,\ 0)$
- $(0,\ 1)$ ✓
- $(-\infty,\ \infty)$

6

If you are using all features of my dataset and you achieve 100% accuracy on my training set using a Multi-layer Neural Network model, but ~70% on validation set, what should you look out for?
(1/1 Points)

- Underfitting
- Nothing, the model is perfect
- Overfitting ✓
- None of the above

7

Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0, 1].
(1/1 Points)

- Sigmoid ✓

- Mode
- Square
- All of the above

8

The effectiveness of an SVM depends upon:
(1/1 Points)

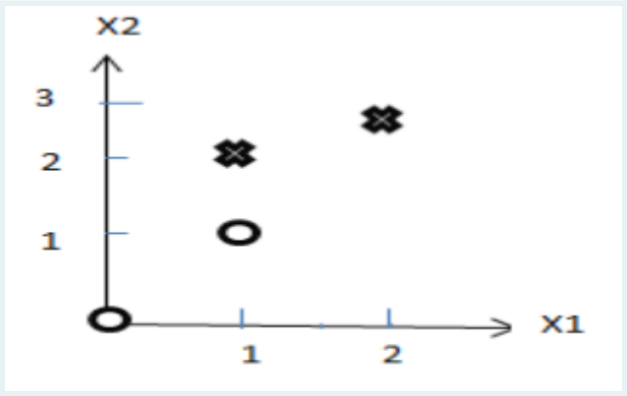
- Selection of Kernel
- Kernel Parameters
- Soft Margin Parameter C
- All of the above ✓

9

How can you prevent a clustering algorithm from getting stuck in bad local optima?
(1/1 Points)

- Set the same seed value for each run
- Use multiple random initializations ✓
- Both A and B
- None of the above

10



Consider the following two-class data set as shown in Figure. Just by visual inspection, find the decision boundary learnt by SVM. [X- Class 1 and O- Class 0].
(2/2 Points)

- $X_1 = 1.5$
- $X_2 = 1.5$ ✓
- $X_1 + X_2 = 1.5$
- *None of these*

11

A feed-forward neural network is said to be fully connected when
(1/1 Points)

- All nodes are connected to each other.
- All nodes at the same layer are connected to each other.
- All nodes at one layer are connected to all nodes in the next higher layer. ✓
- All hidden layer nodes are connected to all output layer nodes.

12

Suppose, you applied a Logistic Regression model on a given data and got a training accuracy X and testing accuracy Y. Now, you want to add a few new features in the same data. Select the option(s) which is correct in such a case. (1/1 Points)

- Training accuracy increases
- Testing accuracy increases or remains the same
- Both (A) and (B) ✓
- None of these

Machine Learning (CS1741) (EL VII) Sessional II

Time: 30 Minutes
Max Marks: 15
Note: Answer all questions. Any missing / misprinted data may be suitably assumed

...

Points: 11/15

1. The XOR function can be realized by
(1/1 Points)

- A Perceptron
- A network of perceptron ✓
- A Hopfield network
- None of these

✗

2. Which of the following is not true about McCulloh-Pitts neurons?
(0/1 Points)

- The interconnections are unidirectional
- The activation is bipolar

- All inhibitory interconnection having same weight
- All exhibitory interconnections having the same weight ✓

3. Which is true for neural networks?
(1/1 Points)

- It has set of nodes and connections
- Each node computes it's weighted input
- Node could be in excited state or non-excited state
- All of the mentioned ✓

4. Consider the following two-class data set as shown in Table below. Plot these data points and just by visual inspection, find the decision boundary learnt by SVM.

X1	X2	t (Class)
0	0	0
1	1	0
1	2	1
2	3	1

(2/2 Points)

- $X_1 + X_2 = 2$
- $X_1 = 1.5$
- $X_2 = 1.5$ ✓
- *None of these*

5. What is perceptron?
(1/1 Points)

- a single layer feed-forward neural network with pre-processing ✓
- an auto-associative neural network

- a double layer auto-associative neural network
- a neural network that contains feedback

6. Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0, 1].
(0/1 Points)

- Sigmoid ✓
- Mode
- Square
- All of the above

7. If you are using all features of my dataset and you achieve 100% accuracy on my training set using a Multi-layer Neural Network model, but ~70% on validation set, what should you look out for?
(1/1 Points)

- Underfitting
- Nothing, the model is perfect
- Overfitting ✓
- None of the above

8. Suppose an input to Max-Pooling layer in a convolutional neural network is given in Table below. The pooling size of neurons in the layer is 3 x 3. What would be the output of this Pooling layer?

3	4	5
4	5	6
5	6	7

(2/2 Points)

- 5
- 6
- 7 ✓
- 8

9. Suppose you have a two-class data set in 2-D space, generated as follows. Obtain the feature transform function) so that a linear SVM can classify the data points correctly.
(0/1 Points)

Positive samples and negative taken from points on the curve $x_1^2 + x_2^2 = 4,$ *and* $x_1^2 + x_2^2 = 12.$

- $\phi(x) = x_1^2 + x_2^2$ ✓
- $\phi(x) = x_1 x_2$
- $\phi(x) = x_1^2 x_2^2$
- *None of these*

10. Which of the following is true about Support Vector Machines (SVMs)?
(0/1 Points)

- Support vectors are used for computing hyperplanes ✓

- It is a method for minimizing the margin to hyperplanes
- Nonlinear problems are handled with mapping inputs to lower-dimensional space
- None of the above

11. A feed-forward neural network is said to be fully connected when
(1/1 Points)

- All nodes are connected to each other.
- All nodes at the same layer are connected to each other.
- All nodes at one layer are connected to all nodes in the next higher layer. ✓
- All hidden layer nodes are connected to all output layer nodes.

12. Assume a simple MLP model with 3 neurons and inputs= 1,2,3. The weights to the input neurons are -4, -5 and 6 respectively. Assume the activation function, f(x) = 3x. What will be the output?
(2/2 Points)

- 5
- 4 ✓
- 6
- 10

Machine Learning (CS1741) (EL VII) Re-Sessional

Time: 30 Minutes Max Marks: 15

Note: Answer all questions. Any missing / misprinted data may be suitably assumed

Points: 13/15

1. Hard margin in SVM allows very low error in classification.
(1/1 Point)

- ◯ True ✓
- ◯ False

2. The effectiveness of an SVM depends upon:
(1/1 Point)

- ◯ Selection of Kernel
- ◯ Kernel Parameters
- ◯ Soft Margin Parameter C
- ◯ All of the above ✓

3. Which of the following is true about Naive Bayes?
(1/1 Point)

- ◯ Assumes that all the features in a dataset are equally important
- ◯ Assumes that all the features in a dataset are independent

- Both A and B ✓
- None of the above

4. Which of the following is true about Unsupervised learning?
(0/1 Point)

- Categorizes training vectors by identifying similarities between them ✗
- Can use the same error functions as supervised learning
- It is collaborative learning methods which often applied between classes
- None of the above ✓

5. In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden output layer and input hidden layer?
(2/2 Points)

- [8 x 5] and [5 x 1]
- [5 x 8] and [5 x 1]
- [8 x 5] and [1 x 5]
- None of these ✓

6. Which of the following is true about k-means algorithm?
(0/1 Point)

- Automatically finds the number of clusters
- Each cluster centre is moved to the mean of data points assigned to it for each iteration ✗

- A too small number of clusters may lead to overfitting
- None of the above ✓

7. Suppose an input to Max-Pooling layer in a convolutional neural network is given in Table below. The pooling size of neurons in the layer is 3 x 3. What would be the output of this Pooling layer?

3	4	5
4	5	6
5	6	7

(1/1 Point)

- 6
- 7 ✓
- 8
- 5

8. The perceptron learning suffers from classifying non-linear data sets and it can't handle _____.

(1/1 Point)

- Overfitting
- Underfitting
- Noise ✓
- None of these

9. Suppose you have a two-class data set in 2-D space, generated as follows. Obtain the feature transform function) so that a linear SVM can classify the data points correctly.
(2/2 Points)

Positive samples and negative taken from points on the curve $x_1^2 + x_2^2 = 4$, *and* $x_1^2 + x_2^2 = 12$.

- $\phi(x) = x_1^2 + x_2^2$ ✓
- $\phi(x) = x_1 x_2$
- $\phi(x) = x_1^2\, x_2^2$
- *None of these*

10. A perceptron with 4 inputs has the weight vector W. The activation function is linear and given by f(yin) = 2yin. If the input vector is X, then find the output of the neuron given the followings.
(2/2 Points)

$W = [1\ 2\ 3\ 4]^T$ *and* $X = [5\ 6\ 7\ 8]^T$

- 140 ✓
- 64
- 128
- 70

11. The polynomial curve fitting often suffers from over-fitting problem.
(1/1 Point)

- True ✓
- False

12. Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0, 1].
(1/1 Point)

- Sigmoid ✓
- Mode
- Square
- All of the above