Midterm Exam Sample and Guidelines

- This is a closed book exam.
- Communicating with anyone else during the exam is considered “cheating”.
- You are not allowed to leave the room during the exam.
- You are not allowed to use earphone during the exam.
- Switch off your cell phone and let is stay in the off mode until the end of the exam.
- You can use a calculator which you need to show to the camera when the exam begins.
- The exam duration is 50 minutes.
- Midterm exam covers all the topics covered so far including Lecture 9, except the “Probabilistic Graphical Models”. Discussions are also part of the material which you need to study for the exam.
- The exam will NOT contain coding questions.
- The exam will NOT ask you to provide mathematical proof.
- The sample questions provided in this guide are just example questions to help you prepare for the exam. It does not mean these are the only questions to expect in the midterm exam. Make sure to study all the lecture videos and discussion videos to prepare for the exam.

1- In classification problems, there may be multiple ways of classifying data items, i.e., a data item may belong to more than one classification category. T/F

2- Which of the following is an example of a flag variable?
   a) Gender: female/male
   b) Weather: clear/rainy/cloudy
   c) Temperature: [21, 80]
   d) a and b

3- Which of the following statements is/are correct?
   a) In machine learning, most of the data is used for testing.
   b) In machine learning, most of the data is used for training.
   c) Training set is used to determine the accuracy of the model.
   d) b and c.

4- Given the following table in the context of machine learning, indicate the target variable, its type, and the type of problem.

Job title	Salary (2014 – 2017)	Job popularity
Software developer	80K , 90K, 100K, 120K	?
University Professor	80K, 85K,88K,95K	?
Bank Employee	65K, 67K, 69K, 71K	?
Mechanical Engineer	77K, 82K,88K,93K	?

a) Salary, numeric, regression problem.
b) Job popularity, numeric, regression problem.
c) Job popularity, categorical , classification problem.
d) Job title, categorical, classification problem.

5- which machine learning technique should be applied to the following problem? "In information retrieval, a search engine needs to find groups of documents that are similar to each other based on important term appearing in them".
a) Clustering
b) Classification
c) Regression
d) Validation

6- Which of the following tasks is an unsupervised learning technique?
a) Clustering
b) Classification
c) Regression
d) All of the above

7- Which of the following methods requires having a training set and test set?
a) Supervised Learning
b) Unsupervised Learning
c) a and b
d) None of the above

8- K-means clustering is an unsupervised technique to partition the dataset into K pre-defined distinct non-overlapping subgroups. T/F

9- Which of the following is NOT an example of a machine learning problem?
a) Optical character recognition: categorize images of handwritten characters by letters represented
b) Face detection: find faces in images
c) Topic spotting: categorize news articles
d) None of the above.

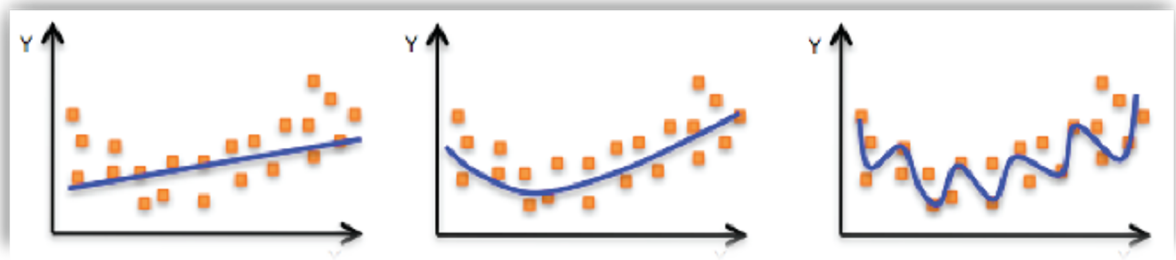
10- In the figure below, each row (sample) indicates information about an animal. The target variable is “label”. Which dependency rule can be inferred from this training set using association rule discovery?

example									label
				train					
aardvark	→	1	1	18	4	22	1	18	11	−
cow	→	3	15	23						+
giraffe	→	7	9	18	1	6	6	5		−
termite	→	20	5	18	13	9	20	5		−
oyster	→	15	25	19	20	5	18			+
dove	→	4	15	22	5					−
spider	→	19	16	9	4	5	18			+
dog	→	4	15	7						+
elephant	→	5	12	5	16	8	1	14	20	+
				test					
rabbit	→	18	1	2	2	9	20			
frog	→	6	18	15	7					
kangaroo	→	11	1	14	7	1	18	15	15	

figure Animal dataset

a) All mammals have positive label.
b) Only insects have negative label.
c) If the 3rd letter of the animal name is an odd number in the alphabet, the label is positive.
d) If the 3rd letter of the animal name is an even number in the alphabet, the label is positive.

11- Indicate the model characteristics depicted in the following figure, from left to right.



a) Overfitting, under-fitting, balanced
b) Under-fitting, balanced, balanced
c) Under-fitting, balanced, overfitting
d) Overfitting, balanced, balanced

12- Association rules are good means to predict sequential dependencies among different events. T/F

13- What does the loss function measure?
a) residual error
b) prediction error
c) model parameters
d) all of the above

14- Why do we use regularization on models?
a) To measure the accuracy of a model
b) To prevent overfitting
c) To train a model
d) All of the above

15- In a 2-layered Neural Network, the perceptron takes an input, calculates the weighted sum of the inputs and weights, and returns 1 if the weighted sum is above a threshold value (T/F)

16- N-fold Cross validation is a method used to prevent overfitting. T/F

17- The following shows a simple linear regression model for a dataset with x1,…,xn attributes. T/F

Answer: it is multiple linear regression because it has multiple features.
If the features are n-degree polynomial, then it is polynomial regression.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

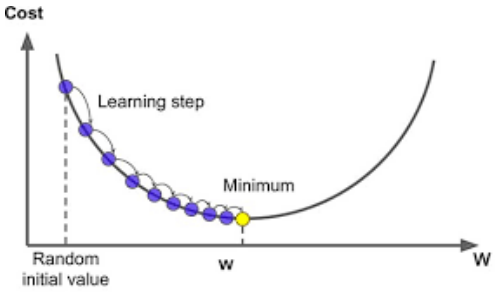
18- When training a model, the main goal is to:
a) Update model coefficients
b) Minimize the error by updating model coefficients
c) Add bias
d) None of the above

19- What is Ordinary Least Squares Method for?
a) Minimize the loss function
b) Maximize the loss function
c) Update the parameters of a model
d) a and c

20- OLS method is used when the relationship between input and output is very complex. T/F

21- Gradient Descent method is used when the relationship between input and output is very complex. T/F

22- For the following loss function convex, if $\frac{\partial RSS}{\partial w_j} > 0$, suggest how the weight shuld be adjusted to near the optimal weight?



Final Answer: Decrease w

23- In the following dataset with input attributes x1,x2 and output variable y, calculate the total prediction error for the following linear regression model:

y = 0.5 + 0.8 x1 - 0.6 x2

X1	X2	y
3	4	0.7
2	5	-0.2
6	1	5

Final answer: 0.31

Solution:
First sample: 0.5+ 0.8(3) - 0.6(4) = 0.5
second sample: 0.5+ 0.8(2) - 0.6(5) = -0.9
third sample: 0.5+ 0.8(6) - 0.6(1)= 4.7

X1	X2	y	y_hat	y_hat - y
3	4	0.7	0.5	-0.2
2	5	-0.2	-0.9	-0.7
6	1	5	4.7	0.3

according to the sum of squares error function E(w):

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

the error is 0.5 (0.04 + 0.49 + 0.09) = 0.31

24- Given a dataset with x1,…, x6 input attributes, how are the terms in a polynomial regression model constructed for this dataset?
   a) Features of the dataset are converted to their higher order polynomial to represent the terms in the model
   b) Features of the dataset are used in the same way as linear regression model
   c) Always two features are used
   d) None of the above

25- Regularization is a method that penalizes model coefficients to reduce overfitting. T/F

26- Lasso is an example of regularization method. T/F

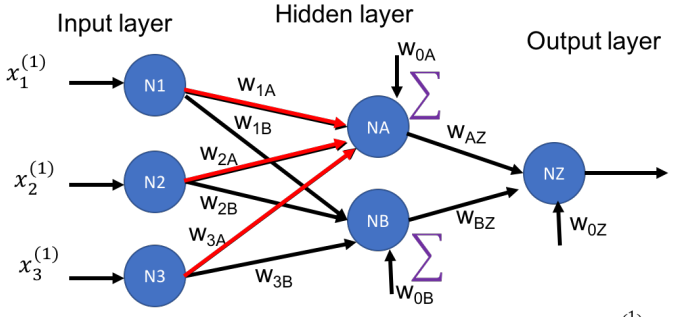
27- MLE uses natural log to optimize the computation cost of the MLE. T/F

28- In the following MLP, calculate the activation of NB in the network if sigmoid function is used as activation function.

Final Answer: 0.8176

Let's show the structure of hidden layer nodes and output layer nodes using the first sample in the dataset D.

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & x_3^{(n)} \end{bmatrix} = \begin{bmatrix} 1 & 0.4 & 0.2 & 0.7 \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & x_3^{(n)} \end{bmatrix}$$



w_{0A} =0.5	w_{0B} =0.7	w_{0Z}=0.5
w_{1A} = 0.6	w_{1B} = 0.9	w_{AZ}=0.9
w_{2A}=0.8	w_{2B}=0.8	w_{BZ}=0.9
w_{3A}=0.6	w_{3B}=0.4	

$$net^{(i)}{}_j = \sum_k w_{kj} x_{kj}^{(i)}$$

$$net^{(1)}{}_A = \omega_{0j} + \omega_{1j} x_{1j}^{(1)} + \omega_{2j} x_{2j}^{(1)} + \omega_{3j} x_{3j}^{(1)} = 1.32$$

Source: Discovering Knowledge in Data D. Larose

19

29- Suppose we have trained a linear regression model y = ax+b where a = 0.5 and b = 1.0, on a set of training data points D = {(1.0,1.6),(1.5,1.5),(3.0,2.4)}. Please calculate the mean squared errors of this model on D.

Final Answer: 0.0275

Solution:
0.5(1) + 1 = 1.5 , 1.5-1.5 = 0.1
…
MSE = ($(0.1)^2$+$(0.25)^2$+ $(0.1)^2$) / 3 = 0.0275

30- what is the name of a 3-layered neural network?

a) Perceptron
b) Multilayer Perceptron
c) Deep Neural Network
d) None of the above

31- What is the popular technique to find the parameters of a Deep neural network?
a) OLS
b) Stochastic gradient descent
c) Mini-batch gradient descent
d) None of the above

32- What is the popular technique to find the parameters of a shallow neural network?
a) OLS
b) Stochastic gradient descent
c) Mini-batch gradient descent
d) None of the above

33- How many output neurons in ANN is needed to perform a binary classification?
a) 1
b) 2
c) 3
d) 4

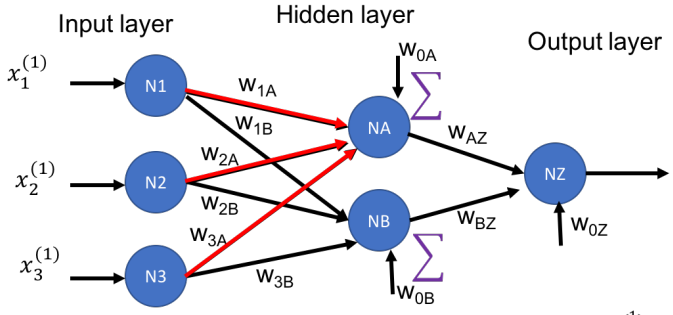
34- how many output neurons in ANN is needed to perform multiclass classification when the output labels are ordered?
a) 1
b) 2
c) 3
d) 4

35- when is 1-of-n output encoding implemented in an Artificial Neural Network?
a) To perform binary classification
b) To perform multiclass classification when output variables are ordered
c) To perform multiclass classification then the output variables are not ordered
d) All the above

36- Compute the output of the following neural network after 1 forward pass when sigmoid function is used as the activation function in all layers:

Let's show the structure of hidden layer nodes and output layer nodes using the first sample in the dataset D.

$$\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & x_3^{(1)} \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & x_3^{(n)} \end{bmatrix} = \begin{bmatrix} 1 & 0.4 & 0.2 & 0.7 \\ \vdots & & \ddots & \vdots \\ 1 & & \cdots & x_3^{(n)} \end{bmatrix}$$



w_{0A} =0.5	w_{0B} =0.7	w_{0Z}=0.5
w_{1A} = 0.6	w_{1B} = 0.9	w_{AZ}=0.9
w_{2A}=0.8	w_{2B}=0.8	w_{BZ}=0.9
w_{3A}=0.6	w_{3B}=0.4	

$$net^{(i)}{}_j = \sum_k w_{kj} x_{kj}^{(i)}$$

$$net^{(1)}{}_A = \omega_{0j} + \omega_{1j} x_{1j}^{(1)} + \omega_{2j} x_{2j}^{(1)} + \omega_{3j} x_{3j}^{(1)} = 1.32$$

Source: Discovering Knowledge in Data D. Larose

Final Answer: 0.8750

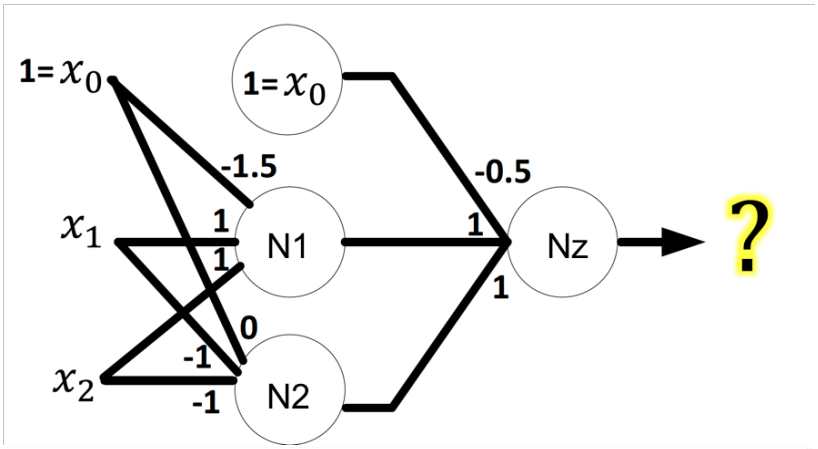
37- In a deep neural network, different activation functions may be used at different layers. T/F

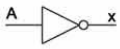
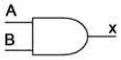
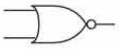
38- A perceptron model can be used to emulate the functionality of AND logical gate. T/F

39- A 2-layered Neural network can be used to emulate the functionality of XOR gate. T/F

40- Identify the logical gate for every neurons N1,N2,Nz respectively in the following figure if threshold is 0:

x_1	x_2	y: $g(x;w)$
0	0	?
0	1	?
1	0	?
1	1	?



Name	NOT	AND	NAND	OR	NOR	XOR	XNOR
Alg. Expr.	$\overline{A}$	AB	$\overline{AB}$	$A+B$	$\overline{A+B}$	$A \oplus B$	$\overline{A \oplus B}$
Symbol							
Truth Table	A X: 0 1; 1 0	B A X: 0 0 0; 0 1 0; 1 0 0; 1 1 1	B A X: 0 0 1; 0 1 1; 1 0 1; 1 1 0	B A X: 0 0 0; 0 1 1; 1 0 1; 1 1 1	B A X: 0 0 1; 0 1 0; 1 0 0; 1 1 0	B A X: 0 0 0; 0 1 1; 1 0 1; 1 1 0	B A X: 0 0 1; 0 1 0; 1 0 0; 1 1 1

Final Answer: AND,NOR, XNOR

Other notes:

- Study all lectures from Lec1~ Lec9. Only Probabilistic graphical models are excluded from the midterm exam.
- Also make sure to study backpropagation , termination criteria in deep learning, mini-batch gradient descent, momentum, hyperparameter tuning, regularization in Deep NN for the exam.