The German University Cairo
Faculty of Media Engineering and Technology
Dr. Aysha Alsafty

November 2017

CSEN1067 Machine Learning Algorithms

Midterm Exam- Second Chance

Winter 2017

Bar code

Instructions: Please Read Carefully Before Proceeding.

1. The allowed time for this exam is 2 hours (120 minutes).
2. Be sure to answer all questions.
3. Please write your solutions in the space provided. If you need more space, please use the back of the sheets.
4. This exam booklet contains 9 pages, including this page.
5. Show your workout and justify your answers in all the questions.

All the best.

Please, do not write anything on this area.

Question	**1**	**2**	**3**	**4**	**5**	**Total**
Maximum Mark	**6**	**10**	**10**	**10**	**4**	**40**
Earned Marks						

Question 1: Introduction to Machine Learning (6 marks)

Design a neural network that simulate a human playing the well-known "snake" game. In this simple version, the snake is trying to maximize points by "eating" prizes that appear on the screen at random locations and time points.

When running (after training is complete), the network will output the action/movement the player should take: up, down, right or left.

The network will be fed the current state of the game (the screen on the user mobile) as input.

The network will be run one time for each frame of the game and will generate one action/movement for the user to make.

Design the neural network by answering the following questions:

a) State the input features and the representation/format they should be fed to the network. [1.5 points]

Two possible answers:

The pixels of the display are flattened as a 1-D input vector. [-0.5 if you did not recognize the need for flattening the 2D pixel map]

OR : The list of pixels occupied by the snake, the coordinates of the prize or -1,-1 if no prize.

b) State the output of the network and its representation/format.

The next player move: up, down, right or left. Represented as 1 out of n classification output (4 output nodes where at most one can fire at a time)

OR: one output neuron, whose output value is 0..1, divided to 4 subranges, each subrange represents a move. (for example 0..0.24 up, 0.25..0.49 down, 0.5..0.74 right, 0.75..1 left)

This is not a very good representation because it imposes artificial mapping between real numeric values and moves.

OR: 3 output nodes, where 4 the possible 8 binary combinations are used to encode moves:

000 up , 010 down, 001, left, 100 right.

This is not an ideal representation, because some binary combination can be generated during the testing or recall phases which will have no interpretation.

c) Does your design describe a regression or a classification neural network?

A classification

For the second solution for part b) above, regression was counted correct.

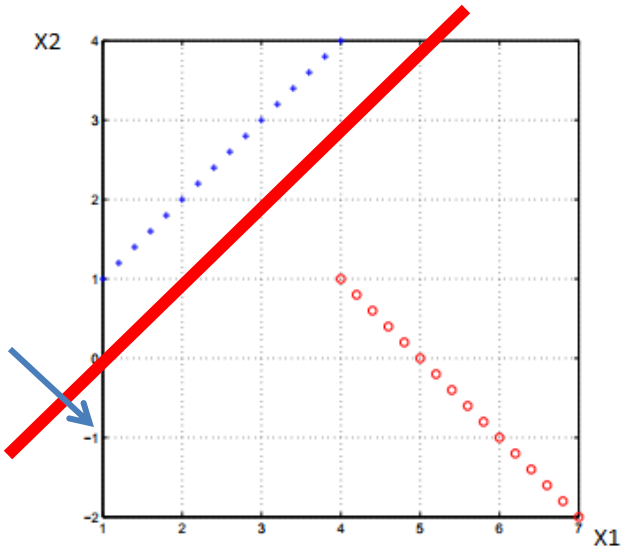
d) What is the activation function that should be used at the output layer? **Justify**.

Softmax. as the player can take at most one move at a time, only one out of the 4 nodes should be considered as having fired.

Question 2: Single Layer Perceptron **(10 marks)**

The dataset in the figure below includes two classes: circles and dots.

2.1. Plot a good decision boundary that separates the two classes on the figure.



2.2. Write the straight line equation for the decision boundary.

Hints:

The decision boundary represents the line where y (the output value) equals zero.

The decision boundary is a function of the input features (X1 and X2).

The decision boundary is the line where Y = 0 (the classifier decision = 0)

Two points on the straight line drawn above are (1,0) and (5,4) → X1 = X2 + 1 → X1-X2-1 = 0 = Y

Y = X1 - X2 - 1

2.3. Based on the straight line equation you devised in 2.2, identify the values for the weight vector and the bias b.

W = [1, -1, -1] OR [1, -1] and bias = -1

2.4. Plot the weight vector on the figure on the previous page. Comment on the geometric relationship between the weight vector and the decision boundary.

Plotted above. The weight vector is perpendicular on the decision boundary. Plotted from the origin and pointing to the [1,-1] point. Any perpendicular vector was considered correct.

Bonus

Prove that the general geometric relationship between the weight vector the decision boundary.

For two points X1 and X2 on the decision boundary , $Y(X1)=Y(X2) = 0$

$Y(X1) = WX1 + b$

$Y(X2) = WX2 + b$

But $Y(X1) = Y(X2) = 0$

$W.X1 + b = W.X2 + b = 0$

$W.X1 – W.X2 = b – b = 0$

$W. (X1 – X2) = 0$

The dot product between the two vectors W and (X1-X2) can be also expressed as

$|W|.|X1 – X2|. \cos \theta = 0$ → (1)

We know that $|W|$ and $|X1 – X2|$ are non zero vectors (not in the general case)

So for the product in (1) to equal 0, then $\cos \theta$ must equal zero

Then θ is 90, and the two vectors are perpendicular. (X1 – X2) is parallel to the decision boundary.

Question 3: MLP Design (10 marks)

3.1. What are the typical values for the learning rate of a neural network?

3.2. Describe the use of an adaptive learning rate while training a neural network.

3.3. Describe when the error calculation and the weight updates are performed in each of:

a) sequential gradient descent.

b) batch gradient descent.

3.4. Which training approach is more appropriate for use on a training dataset with many redundant points; batch or stochastic gradient descent? **Justify**.

Question 4: MLP Training (10 marks)

4.1. Define underfitting. How would underfitting affect the accuracy/loss of a neural network? **Justify**.

Underfitting occurs when a model (linear or nonlinear) fails to fit a considerable ratios of the data points representing the phenomenon we are the model is trying to represent. It can be observed in neural networks when the training was not sufficient, or the network architecture is too simple to represent a complex dataset.

Underfitting means that maximum accuracy of the network is too low, in other words, the loss remains high.

Justification: as the model is too simple or too inaccurate to represent the underlying phenomenon, many of the data points will be misclassified or mis-evaluation in a regression model. Hence the accuracy of the model suffers.

4.2. If a neural network is underfitting the dataset, **explain and justify** the appropriate measure to take in each of the two cases below:

a) if underfitting is observed early on in the training

it is normal for a neural network to underfit a dataset early on in the training. Hence, no measure should be taken, except for continuing the training as usual and observing that the accuracy is increasing at an acceptable rate.

b) if underfitting is observed after training is complete.

This could be an indication that the model / network architecture was too simple to represent the dataset at hand, in which case the model should be re-considered and the training must be restarted with the new model.

If the model/ neural network architecture are thought to be complex enough to represent the dataset, then it is possible that the model is udnerfitting because the training was not thorough enough, hence the network converged to a local minimum with a very high error value. Training can be restarted with the same model, however a smaller learning rate can be used to make sure that the error surface was well-explored and that the model did not converge prematurely.

4.3. Standardization is preprocessing the training data so that it has a zero mean (mean cancellation) and a unit standard deviation and variance.

What are the possible side effects of training a multi layer perceptron over a non-standardized dataset.

4.4. Compare the test dataset and the validation dataset with respect to:

a) The objective of their use

b) The stage/time they are applied to the neural network.

Question 5: Backpropagation **(4 marks)**

5.1. Show the steps for deriving the gradient of the error function for a multilayer perceptron with respect to a weight in the outer layer (the layer connecting the hidden layer to the output layer).