Midterm

Wednesday, October 20, 2021 1:16 AM

EE628WS Midterm

Q1. Which of the following statement(s) correctly represents a concept of neuron?

(a) A neuron has a single input and a single output only
(b) A neuron has multiple inputs but a single output only
(c) A neuron has a single input but multiple outputs
(d) A neuron has multiple inputs and multiple outputs
(e) All of the above statements are valid

Neurons can have one or more inputs & one or more outputs.

Q2. (a) • Once a dataset's dimensionality has been reduced, is it possible to reverse the operation? If so, how? If not, why?

(b) • Can principal component analysis (PCA) can be used to reduce the dimensionality of a highly nonlinear dataset? Justify your answer.

(c) • Does it make sense to chain two different dimensionality reduction algorithm?

(a) Once a dataset's dimensionality has been reduced, depending on the method used for the dimension reduction, the operation can potentially be reversed, but usually only partially. Some dimension reduction techniques, such as PCA, can be easily reversed (while others may not as easily be reversed), there will still data that has been others may not as easily be reversed, there will still data that has been lost during the reduction, meaning that it will be impossible to perfectly reverse the operation.

(b) While PCA can be used for dimension reduction on a highly non-linear dataset, it will likely not produce good or meaningful results. PCA works best on linear data. If PCA is done on non-linear data, you will end up with distorted results.

(c) It is indeed a common technique to chain more than one dimension reduction algorithm. Take, for example, Fisher's Linear Discriminant & PCA. Both are used for dimension reduction but using different methods to achieve different results. For FLD to work different results. For FLD to work well, the number of samples N has to be greater or equal to the data dimensionality D. When D > N, though, which is common in vision problems like facial recognition, it could be useful to use PCA first to reduce D, & then use FLD to reduce the data dimensionality even further + maximize the separability of the classes.

Q3. In a neural network, knowing the weight and bias of each neuron is the most important step. If you can somehow get the correct value of weight and bias for each neuron, you can approximate any function. What would be the best way to approach this?

(a) Assign random values and pray to God they are correct

(b) Search every possible combination of weights and biases till you get the best value

(c) Iteratively check that after assigning a value how far you are from the best values, and slightly change the assigned values values to make them better

(d) None of these

This is essentially gradient descent.

Q4. • For a fully-connected deep network with one hidden layer, increasing the number of hidden units should have what effect on bias and variance? Explain briefly.

Increasing the number of hidden units

Increasing the number of hidden units (neurons) will make the model more complex. In doing so it will decrease the bias & as a result decrease under-fitting. However, there is a tradeoff. The more complex the model, the higher the variance & the higher the chance of over-fitting. You need to find the right amount of hidden units to give the minimum bias & variance.

Q5. What are the steps for using a gradient descent algorithm?

1. Calculate error between the actual value and the predicted value
2. Reiterate until you find the best weights of network
3. Pass an input through the network and get values from output layer
4. Initialize random weight and bias
5. Go to each neurons which contributes to the error and change its respective values to reduce the error

(a) 1, 2, 3, 4, 5

(b) 5, 4, 3, 2, 1

(c) 3, 2, 1, 5, 4

(d) 4, 3, 1, 5, 2

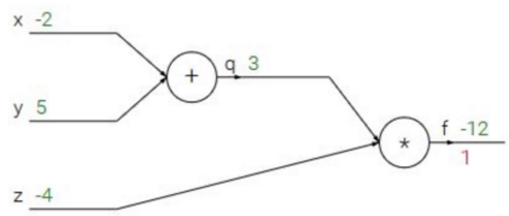
Q.6 Choose the true statements about text tokens:

(a) Stemming can be done with heuristic rules

~~(b) Lemmatization is always better than stemming~~

(c) Lemmatization needs more storage than stemming to work

(d) A model without stemming/lemmatization can be the best



Q7. Suppose you have inputs as x, y, and z with values -2, 5, and -4 respectively. You have a neuron "q" and neuron "f" with functions:

$$q = x + y$$

$$f = q * z$$

Graphical representation of the functions is as above:

What is the gradient of f with respect to x, y, and z?

(HINT: To calculate gradient, you must find (df/dx), $(df/dy) and (df/dz)$)

(a) (-3, 4, 4)

(b) (4, 4, 3)

(c) (-4, -4, 3)

(d) (3, -4, -4)

$f = (x+y) \cdot z = xz + yz$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} = \begin{bmatrix} z \\ z \\ x+y \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \\ -2+5 \end{bmatrix} = \boxed{\begin{bmatrix} -4 \\ -4 \\ 3 \end{bmatrix}}$$

Plugging in x, y, + z

Q8. Which of the following gives non-linearity to a neural network?

(a) Stochastic Gradient Descent

(b) Rectified Linear Unit

(c) Convolution function

(d) None of the above

RELU is one of the most common activation functions.

Q9. Which gradient technique is more advantageous when the data is too big to handle in RAM simultaneously?

(a) Full Batch Gradient Descent

(b) Stochastic Gradient Descent

Stochastic gradient descent uses a single data-point per iteration. Full batch uses all the data.

Q10. Which of the following is true about model capacity (where model capacity means the ability of neural network to approximate complex functions) ?

(a) As number of hidden layers increase, model capacity increases

(b) As dropout ratio increases, model capacity increases

(c) As learning rate increases, model capacity increases

(d) None of these

Model capacity is similar to complexity.

Tim Demetriades