CS503 | Machine Learning | Quiz-2 [2022]

Total points 16.5/25

+ There are a total of 9 questions.
+ There is 1 subjective question whose answer is to be written on paper and the PDF of answer script is to be uploaded in this form itself. Naming convention YourName_RollNo.pdf
+ The total time for the quiz is 45 min.
+ This quiz is for 25 marks.
+ Keep a pen and paper for rough work.
+ Some responses are for information and not graded.

All the Best!!

Name *

Ashish Sharma

Department *

Mathematics and Computing

Roll No *

2019MCB1213

Honour Code/Declaration: *

I declare that I will be fair and honest during the online examination. I will neither give nor receive unauthorized aid on any question in this quiz/exam. I also understand that it is my obligation to report violations of this honour code to the instructor. If I have been found/discovered in violation of the honour code then I accept that I will be liable to receive the 'Fail' grade in this course.

- (●) I Agree

Choose the most appropriate response:

✓ Suppose the number of nodes in the input layer is 5 and in the hidden layer it is 10. The maximum number of connections from the input layer to the hidden layer would be- 2/2

- ○ More than 50
- ○ Less than 50
- ◉ 50 ✓
- ○ 5^10

✓ A perceptron model can represent which of the following functions: 2/2

- ☑ A ∧ B ✓
- ☐ (A ∧ ¬B) ∨ (¬A ∧ B)
- ☑ A ∧ B ∧ ¬C ✓
- ☐ None of the above

✓ Adding the sum of all weight parameters to the Error function E helps in: (Choose all that apply) 2/2

- ☑ Improving the generalization accuracy of the model ✓
- ☐ Improving the training accuracy of the model over the given dataset
- ☑ Forces gradient descent to find weight vectors with small magnitude ✓
- ☑ Avoids overfitting the datatset ✓

✗ Which of the following are true about gradient descent? (select all statements that are true.) 1/2

- [ ] After each iteration, we modify the weight vector in the direction of the gradient.
- [ ] We have to choose a non-variable learning rate.
- [x] After each iteration, we modify the weight vector in the direction of the negative gradient. ✓
- [ ] In the gradient descent algorithm each update of the weight vector depends on all the training examples.

Correct answer

- [x] After each iteration, we modify the weight vector in the direction of the negative gradient.
- [x] In the gradient descent algorithm each update of the weight vector depends on all the training examples.

✓ Parameter Sharing is a feature of 2/2

- [ ] MLP
- [x] CNN ✗
- [x] RNN ✗
- [ ] None of the above

No correct answers

✓ Consider a 10-class prediction problem. Which of the following configuration must be used in the output layer of an ANN? 2/2

- (●) Activation function = softmax, Error function = cross entropy ✓
- ( ) Activation function = sigmoid, Error function = cross entropy
- ( ) Activation function = softmax, Error function = mean squared error
- ( ) Activation function = sigmoid, Error function = mean squared error

Given a dataset having 1000 instances. The target label is binary while the input vector has 5 features. Assuming you train an ANN on this dataset for 50 epochs, how many weight updates will be applied in case of :

✗ Mini-batch gradient descent. with batch size of 100 0/2

5000 ✗

Correct answer

500

✗ Gradient Descent 0/2

500 ✗

Correct answer

50

✗ Stochastic Gradient Descent 0/2

500000 ✗

Correct answer

50000

Solve

✓ Consider the following : (Write your answer as comma separated values in order as [w_1, w_2, w_3]) 3/3

Assume you have three training instances in dataset D, each with 3 features:
Instance 1: [1, 0 , 1] Target Label: 1
Instance 2: [0, 1 , 1] Target Label: 0
Instance 3: [1, 0 , 0] Target Label: 1

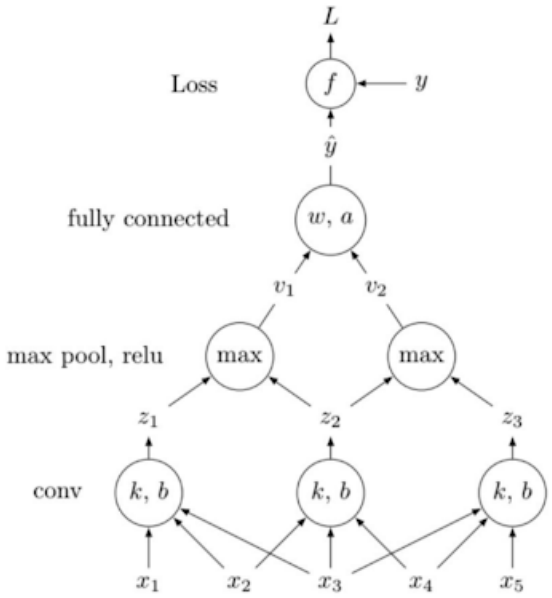
Consider the initial weight vector as w = [0, 0, 0] (one weight parameter for each feature). Using gradient descent with Least Squared error $E(w) = \frac{1}{2}\sum_{d \in D}(y_d - w^T x_d)^2$ and learning rate $\eta = 1$, calculate the first update vector ∇w to be added to w = [0, 0, 0].

[2, 0, 1] ✓

Subjective: Answer this question on paper and upload using the upload button at the bottom of this form

2.5/4

Consider the following 1-dimensional ConvNet, where all variables are scalars:



$$L = \frac{1}{2}(y - \hat{y})^2$$

$$\hat{y} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + a$$

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \max\{z_1, z_2, 0\} \\ \max\{z_2, z_3, 0\} \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 & 0 & 0 \\ 0 & k_1 & k_2 & k_3 & 0 \\ 0 & 0 & k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix}$$

Given, $\frac{\partial L}{\partial v_1} = \delta_1$ and $\frac{\partial L}{\partial v_2} = \delta_2$

Show the calculation for:

$$\frac{\partial L}{\partial z_1}, \frac{\partial L}{\partial z_2}, \frac{\partial L}{\partial z_3}$$

- (●) Attempted
- ( ) Not Attempted

No correct answers

Upload your pdf

PDF ASHISH_SHARM...