Quiz Submissions - Quiz 1

WENXIN ZHANG (username: wz2164)

Attempt 1

Written: Sep 21, 2021 2:17 PM - Sep 23, 2021 11:33 PM

Submission View

Your quiz has been submitted successfully.

Question 1 **1 / 1 point**

Explain about **model parametrization** in your own words

There are inherent uncertainties within the data and model parameterization is to incorporate these uncertainties. To conduct model parameterization is to figure out p(y|x), the distribution of the output y given the input x, which is proportional to

$$e^{F(x,y)}$$

and is chosen for its benefit in normalization. Sometimes, we also have to consider the latent variables, where we regard the distribution as p(y,z|x) that is proportional to

$$e^{F(x,y,z)}$$

. It indicates that there exists another non-negligible factor that would effect our final output.

Question 2 **1 / 1 point**

Given two distributions p and q, write down the cross-entropy of q relative to p over a given set,

$$H(p,q) =$$

$$H(p,q) = \sum_{x \in X} -p(x) * log(q(x))$$

Question 3 **1 / 1 point**

Show cross-entropy is equivalent to negative log likelihood for a classification problem with one-hot labels

In the classification problem with one-hot labels, the likelihood of a single prediction with true class j can be represented as:

$$\prod_{i=1}^{N} \hat{y_i}^{y_i} = \hat{y_i}^{y_i}$$

. The negative log likelihood is shown as

$$\sum_{i=1}^{N} y_i log(\hat{y}_i) = y_i log(\hat{y}_i)$$

accordingly.

Considering cross entropy, the likelihood of a single prediction can be written as

$$\sum_{x \in X} -p(x) * log(q(x))$$

, where p(x) stands for the probability of our target class of x and q(x) represents the probability of our predicted class of x. In the case of one-hot labels, this can be rewritten as

$$\sum_{i=1}^{N} -y_i * log(\hat{y}_i) = -y_i * log(\hat{y}_i)$$

, since only the component of the right class could be non-zero.

Overall, it's proved that the cross-entropy is equivalent to NLL for a classification problem with one-hot labels.

Question 4 **1 / 1 point**

If we have a vocabulary of 2000 tokens and a token embedding space with a dimension of 300, please select the right shape of the embedding matrix from below:

- ◯ 2000×2000
- ◯ 2000×300
- ◯ 300×300
- ◯ 300×2000

Question 5 **1 / 1 point**

For the bag-of-words model we used in lab2 for multi-class classification, does increasing the embedding dimension always helps to improve validation performance?

- ◯ True
- ◯ False

Question 6 **1 / 1 point**

Regarding the way a sentence or a document is represented in the Bag-of-words (BoW) model, what kind of information is missed in order to accurately convey the message of the text? Select the ones that are most reasonable for you.

- [ ] word order
- [ ] word occurrence / frequency

Question 7 **1 / 1 point**

Given a computed loss variable in PyTorch, obtained as below:

criterion = nn.CrossEntropyLoss()

output = model(input)

loss = criterion(output, label)

Write down the expression to perform backpropagation.

Answer: loss.backward()

Question 8 **1 / 1 point**

In lab 2, after the backpropagation, we perform the updating step with Pytorch SGD Optimizer (*optimizer = torch.optim.SGD(model.parameters(), lr=learning_rate)*) by calling *optimizer.step()*. Here you can find the functional implementation of SGD in PyTorch (https://github.com/pytorch/pytorch/blob/master/torch/optim/_functional.py#L146).

Based on the above code, suppose that we have **weight_decay=0.01, momentum=0, lr=0.1, dampening=0,** for a given parameter **p** and its gradient **g**, select the corresponding update on **p** from the following options:

- ○ $$p = p + 0.1 \times g - 0.001 \times p$$
- ○ $$p = p - 0.1 \times g$$
- ○ $$p = p - 0.01 \times g + 0.1 \times p$$
- ○ $$p = p - 0.001 \times p - 0.01 \times g$$
- ○ $$p = p - 0.1 \times g - 0.001 \times p$$

Question 9 **0.1 / 0.1 points**

How long did you spend to complete this quiz (in hours)?

Nearly 0.8 Hour.

Done