

6406533484213. ✖ $\sigma(Uh_{t-1} + Wx_t)$

6406533484214. ✖ $\sigma(Uh_{t-1}) + \sigma(Wx_t)$

6406533484215. ✖ $\sigma(Wh_{t-1} + UEx_t)$

Question Number : 173 Question Id : 6406531030045 Question Type : MSQ

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

At training time, consider a training sample, "this is good", where each word is a token. At the last time step (while processing "good"), the word probabilities output by the RNN for these three tokens are:

$$p_{\text{this}} = 0.001$$

$$p_{\text{is}} = 0.001$$

$$p_{\text{good}} = 0.4$$

Which of the following are true?

Options :

The total loss for this sample is

$$-\frac{1}{3} [\log(0.001) + \log(0.001) + \log(0.4)]$$

6406533484216. ✖

6406533484217. ✖ The loss for this time step is $-\log(0.001)$

6406533484218. ✔ The loss for this time step is $-\log(0.4)$

If we predict (sample) the word at this time step using a greedy strategy, the output

6406533484219. ✔ could be something different from "good".

DLP

Section Id :

64065374264

Section Number :

10

Section type :

Online

Mandatory or Optional :

Mandatory

Number of Questions :	16
Number of Questions to be attempted :	16
Section Marks :	50
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	No
Section Maximum Duration :	0
Section Minimum Duration :	0
Section Time In :	Minutes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653157145
Question Shuffling Allowed :	No

Question Number : 174 Question Id : 6406531030046 Question Type : MCQ

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DEGREE LEVEL : DEEP LEARNING PRACTICE (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?

CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406533484220. ✓ YES

6406533484221. ✗ NO

Sub-Section Number : 2

Sub-Section Id : 640653157146

Question Shuffling Allowed : Yes

Question Number : 175 Question Id : 6406531030047 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

Which of the following commands can be used to convert a 44.1KHz .mp3 file to a 16KHz .wav file using ffmpeg?

Options :

6406533484222. ✗ `ffmpeg -i input.mp3 -ar 44100 output.wav`

✓ 6406533484223. ✓

input ⇒ .mp3
output ⇒ .wav

ffmpeg -i input.mp3 -ar 16000 output.wav
 -ar sets the audio sample rate $\Rightarrow 44.1 \text{ kHz}$
 6406533484224. ✖ ffmpeg -i input.wav -ac 16000 output.mp3 \downarrow 16 kHz
 6406533484225. ✖ ffmpeg -az 16000 -i input.mp3 output.wav

Question Number : 176 Question Id : 6406531030052 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

In the following code snippet, what will be the shape of x after performing mean across the 3rd dimension, `squeeze(0)`, and `unsqueeze(2)`?

```
x = torch.rand(1, 3, 4, 5)
x = x.mean(dim=2).squeeze(0).unsqueeze(2)
```

Options :

6406533484242. ✖ (2, 3, 4, 1)

6406533484243. ✖ (1, 3, 5, 1)

6406533484244. ✖ (3, 5, 1)

6406533484245. ✔ (3, 1, 1, 5)

Handwritten notes for Question 176:
 (1, 3, 4, 5) → 3rd & 4th dim.
 ↓
 (1, 3, 5) → dim = 2
 ↓
 (3, 5) → x.squeeze(0)
 ↓
 (3, 5, 1) → x.unsqueeze(2)

Question Number : 177 Question Id : 6406531030055 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

For a 1D convolution, given the following values:

- Input size (x): 50
- In channels: 16
- Kernel size: 3
- Padding: 1
- Dilation: 1
- Stride: 2
- Out channels: 32

$$\text{Output} = \frac{\text{input} + 2 \times \text{Pad} - \text{dila}^n \times (K - 1)}{\text{Stride}} + 1$$

$$= \frac{50 + 2 - 1 \times 2}{2} + 1$$

$$= \frac{49}{2} + 1 = 24.5 + 1 = 25.5 \approx 25$$

Handwritten note: 32 x 25

What is the shape of the output of the convolutional layer?

Options :

6406533484254. ✔ 32 × 25

6406533484255. ✖ 32 × 26

6406533484256. ✖ 16 × 32 × 50

6406533484257. ✖ 50 × 3

Question Number : 178 Question Id : 6406531030057 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

If an audio is of duration 10 seconds and sample rate is 48KHz, what happens to the length of audio in samples if the audio is downsampled to 16KHz?

Options :

6406533484262. ✖ Remains the same

6406533484263. ✔ Becomes $\frac{1}{3}$ of the original length

6406533484264. ✖ Triples in length

6406533484265. ✖ Halves in length

$$\begin{array}{ll} \text{duration} & \text{SR} \\ 10 \text{ se} & 48 \\ \downarrow \frac{1}{3} & \downarrow \frac{1}{3} \\ \frac{1}{3} \times 10 & 16 \end{array}$$

Question Number : 179 Question Id : 6406531030060 Question Type : MCQ

Correct Marks : 4

Question Label : Multiple Choice Question

If the current learning rate is 0.002 and the maximum learning rate to be achieved in 1000 steps is 0.004, what should be the step size for a linear learning rate scheduler?

Options :

6406533484274. ✖ 0.0002

6406533484275. ✖ 0.00002

6406533484276. ✖ 0.0000002

6406533484277. ✔ 0.000002

Sub-Section Number : 6

Sub-Section Id :

Question Shuffling Allowed :

$$\begin{array}{l} \text{current } \eta = 0.002 \\ \text{max } \eta = 0.004 \\ \text{steps} = 1000 \end{array}$$

$$\text{step size} = \frac{\text{max} - \text{current}}{\text{steps}}$$

3

640653157147

Yes

$$= 2 \times 10^{-6}$$

Question Number : 180 Question Id : 6406531030053 Question Type : MSQ

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

What is the correct order of training commands?

Options : doubtful

6406533484246. ✔ `loss.backward(); optimizer.step(); optimizer.zero_grad()`

6406533484247. ✖ `optimizer.zero_grad(); optimizer.step(); loss.backward()`

6406533484248. ✔ `optimizer.zero_grad(); loss.backward(); optimizer.step()`

6406533484249. ✖ `loss.backward(); optimizer.zero_grad(); optimizer.step()`

Sub-Section Number :

4

Sub-Section Id :

640653157148

Question Shuffling Allowed :

Yes

Question Number : 181 Question Id : 6406531030048 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

To quantize the Whisper model to 8-bit integer during loading in Hugging Face, which of the following arguments should be passed?

Options :

6406533484226. ✖ `quantization_dtype="int8"`

6406533484227. ✔ `compute_dtype="int8"`

6406533484228. ✖ `model_dtype="int8"`

6406533484229. ✖ `loading_dtype="int8"`

see you can do
load_in_8bit=True

Question Number : 182 Question Id : 6406531030049 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following models can be used to obtain segments for speaker diarization?

Options :

6406533484230. ✔ `Whisper`

6406533484231. ✖ `Wav2Vec`

6406533484232. ✖ `Speecht5`

6406533484233. ✖ `GPT-3`

Whisper is a pre-trained
model that makes
segments and aug. cuts
the embeddings.

Question Number : 183 Question Id : 6406531030050 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following arguments should be passed during dataset loading to avoid loading the full dataset at once?

Options :

6406533484234. ✖ `lazy=False`

6406533484235. ✔ `streaming=True`

6406533484236. ✖ buffering=False

6406533484237. ✖ shard=True

Question Number : 184 Question Id : 6406531030051 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following loss functions can be used for speaker classification?

Options :

6406533484238. ✖ Mean Squared Error (MSE)

6406533484239. ✔ Cross-Entropy Loss

6406533484240. ✖ KL Divergence Loss

6406533484241. ✖ Smooth L1 Loss

others are -
AAM softmax
Contrastive
Center
Info NCE

Question Number : 185 Question Id : 6406531030054 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

If a dataset has 25600 samples, batch size is 100, each epoch processes the whole dataset only once, and gradient accumulation is set to 32, how many updates does the model make?

Options :

6406533484250. ✖ 16

6406533484251. ✖ 256

6406533484252. ✖ 32

6406533484253. ✔ 8

$$\text{batch per epoch} = \frac{25600}{100} = 256$$
$$\text{updates} = \frac{256}{32} = 8$$

Question Number : 186 Question Id : 6406531030056 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following functions can be used to replace certain values in a PyTorch tensor?

Options :

6406533484258. ✔ torch.masked_fill

6406533484259. ✖ torch.scatter

used in
masking the speech
embeddings

6406533484260. ✖ torch.fill

6406533484261. ✖ torch.expand

Question Number : 187 Question Id : 6406531030058 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following trainers is commonly used for ASR sequence-to-sequence training?

Options :

6406533484266. ✖ ASRTrainer

6406533484267. ✔ Seq2SeqTrainer

6406533484268. ✖ LanguageModelTrainer

6406533484269. ✖ ClassificationTrainer

Question Number : 188 Question Id : 6406531030059 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

What type of model is HiFi-GAN? → uses a discriminator

Options :

6406533484270. ✖ Text-to-Text model

6406533484271. ✔ Vocoder

6406533484272. ✖ Decoder

6406533484273. ✖ Classifier

it generates from CNN
that was getting the
Mel spectrograms to audio
waveform.

Question Number : 189 Question Id : 6406531030061 Question Type : MCQ

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following should be done when a CUDA Out of Memory error is encountered during training?

Options :

6406533484278. ✖ Increase batch size

6406533484279. ✔ Decrease batch size

6406533484280. ✖ Use double precision for calculations

6406533484281. ✖ Increase accum_grad

batch size ↓, reduces
the amount of
data processed in 1
go.