Quizizz
Quiz-4

Date: Fri Dec 03 2021 - 2:28 pm

94	17	1
Accuracy % (correct / total)	Number of Questions	Number of Players

Players

Rank	Player Name	Avg. Time	Score	Accuracy	Correct / Total	
1	Konda Mopuri (Konda Mopuri)	119 secs	10890	● 94%	16 / 17	▬▬▬▬▬

Quizizz
Quiz Name:Quiz-4
Student:Konda Mopuri (Konda Mopuri)
Date:Fri Dec 03 2021 - 2:28 pm

94%
Accuracy % (correct / total)

17	16	1	0
Total Questions	✔ Correct	✖ Incorrect	⊘ Unattempted

Questions

No.	Question	Time Taken	Score	Response
1	In the word2vec learning, you estimate P(t|c), where t is the target word and c is a context word. How are t and c chosen from the training set? Pick all the correct answers. [1 Mark]	19 secs	870	✔ c and t are chosen to be nearby words
2	In NLP, words represented as vectors are called as Neural Word Embeddings. True or False? [1 Mark]	7 secs	950	✔ True
3	CBoW configuration of the word2vec doesn't capture the language structure because of its bag of words nature. Support or contradict in about 2 sentences. [1+2 marks]	215 secs	600	No, it does capture the structure in the language because the embeddings are learned while predicting the missing word from the given context. One should not get mislead by the BoW part in the name, which is from the way the context is summed destroying the word ordering. However, the CBoW configuration results in meaningful representations that reflect the language structure.

No.	Question	Time Taken	Score	Response
4	What is the disadvantage of one-hot vector encoding of words apart from memory and computational overhead? Answer in 1-2 sentence(s). [1 Mark]	132 secs	600	It can't capture the structure in the language
5	Given embedding vectors for two unknown words, how can we find their similarity? Answer in about 1-2 sentence(s). [1 Mark]	47 secs	600	We can find distance between those vectors such as cosine similarity/distance or dot product.
6	Are GANs an implicit or explicit density models? Explain your answer in in about 2-3 sentences ? [1+2 marks]	170 secs	600	GANs are implicit density models. Since there is no parametric assumption on the high-dimensional density under investigation, instead GANs learn to sample data the unknown density.
7	What is the mode collapse problem in case of GANs? Describe in about 2-3 sentences [2 Marks]	192 secs	600	While learning the complex high-dim distribution, GAN may end up learning only a subset of the data variety. In other words, if the actual unknown data distribution can be thought of having multiple modes, GAN fails to learn all of them. Result of this is producing fake data which is similar to only a (strict) subset of the full training data.
8	What is the important layer used in the Generator architecture of the DC GAN? [1 Mark]	51 secs	0	✖ DCGAN generator has to create a 2D (image like) input from a vector input. Hence it uses transposed convolution (or, deconvolution) layers.
9	For GAN evaluation via Inception score we do not need to train a classifier on the training data. [1 Mark]	7 secs	910	✔ False
10	How does the encoder of a VAE differ from that of an autoencoder? (describe in about 2 sentences) [2 Marks]	89 secs	600	Encoder of a VAE outputs the mean and variance of the distribution in the latent space for embedding an input. On the other hand the autoencoder computes a single embedding for each input.

No.	Question	Time Taken	Score	Response
11	Describe in 2-3 sentences what reparameterization is in the context of VAEs and why it is needed. [3 Marks]	178 secs	600	After encoder predicts the mean and covariance in the embedding space, decoder has to sample using these and start the decoding process. This causes a glitch in the computational graph fro conducting backpropagation. To mitigate this, sampling operation is made into an analytic form using random noise along with the predicted mean and variance in the latent space. This is called reparameterization trick that helps the smooth learning.
12	What is memorization issue in the context of VAES? Describe in 2-3 sentences how one can avoid it. [3 marks]	492 secs	600	While learning the projections (or, encodings) for the inputs, similar inputs may get different representations in the latent space, causing the embedding space of a set of related inputs to be discrete. This is known as the problem of input memorization. This causes the VAE to learn representations that are not very meaningful. Ideally we want the system to learn continuous latent representations. To avoid this, one may force the distribution of the latent representation [q(z/x)] to be close to a standard distribution such as Gaussian. This makes the related embeddings to be closer in latent space and thus making it smooth.
13	For visualizing the sensitive regions in the input image with respect to a trained CNN classifier, we compute the derivative of the following quantity with respect to the input. [1 Mark]	6 secs	960	✔ Any of the above
14	A Neural network trained to perform 10-way classification accepts an input tensor of shape 64 X 64 X 3. What is the shape of gradients for (i) classification loss, and (ii) probability predicted for a chosen class, computed with respect to the input tensor? [1 mark]	29 secs	600	(i) 64 X 64 X 3 (ii) 64 X 64 X 3

No.	Question	Time Taken	Score	Response
15	What is covariate shift in DNNs? Why is it undesirable? Answer in about 2-3 sentences. [3 Marks]	233 secs	600	During the course of training, the distribution of activations at any specific layer tends to change (or drift) with time. This is because the weights and biases of the neural network change during the training resulting in change of the activations. Hence the succeeding layers need to adapt to this change while trying to optimize for loss minimization. This becomes an additional responsibility on the network parameters and may cause optimization to be ineffective. If we can avoid the activation drift, optimization can then be performed better.
16	Are there any learnable parameters in a batch normalization layer? If yes, what are they? If no, justify your answer. [2 Marks]	91 secs	600	There are two learnable parameters gamma and beta. The mean and variance computed for every mini-batch are not learned, instead they are computed from the current mini-batch samples. Only beta and gamma are learned in the gradient descent framework using the backpropagation algorithm.
17	What are the disadvantages of conducting occlusion sensitivity of a trained CNN for finding the important portions in the input image? (Comma separate the list of phrases) [1 Mark]	61 secs	600	Need to investigate different occlusion size, computationally expensive, etc.