Deep Learning 2022: PRACTICE EXAM

Name and student ID: _____
DATE: _____

1 Answer sheet

	A	B	C	D
1		x		
2	x			
3			x	
4	x			
5				x
6	x			
7		x		
8	x			
9				x
10			x	
SUM:				

2 Questions

1. We build a two-layer feed-forward network, and don't include any activation function. Which is correct?
   A. This network suffers from vanishing gradients because we didn't include sigmoid activations.
   B. This network can also be represented by a single linear layer.
   C. This network can also be represented by a linear function.
   D. This network can also be represented by a multi-layer network with activation functions.

2. Backpropagation is based on the chain rule of calculus.

   If we apply the chain rule to a simple feedforward neural network, the factors we get require things like the derivative of a function with matrix in put and vector output. Such derivates are most naturally expressed as 3-tensors, which require a lot of memory to store.

   Why is it that we can still apply backpropagation to a feedforward network efficiently?
   A. We only ever focus only on the derivative of the loss, and accumulate the product of the chain rule from the loss to the inputs.
   B. We unroll the network in the time dimension.
   C. We flatten all matrices into vectors, making all modules vector-to-vector functions.
   D. We focus only on the scalar view of backpropgation, looping over the individual elements of the tensors.

3. It is known that GANs tend to generate images of high quality, however, they can suffer from serious problems. Which of the following scenarios is one of such problems:
   A. For colorful images, GANs can skip some channels (e.g., generate only gray images).
   B. For images of faces, GANs can generate cars instead.
   C. For handwritten digits, GANs can generate only a subset of images (e.g., only digits 1, 4, and 7).
   D. GANs do not suffer from any problem.

4. We build a sequence classifier by stacking a number of RNN layers followed by a global pooling operation. This results in a single vector which we then project down to the number of classes.

   Why is global pooling preferable to just picking the last element of the output sequence to represent the whole?
   A. It means every token in the input has an equally short path to the output.
   B. It is cheaper to compute.
   C. It reduces the output resolution so we can add more channels.
   D. It mean we will not need an activation in the final layer.

5. Please indicate components of GANs:
   A. Painter, expert, Pablo Picasso.
   B. Critic, discriminator, generator, Wasserstein loss
   C. Discriminator, generator, integrator.
   D. Discriminator, generator, adversarial loss

6. Which one is true?
   A. Self-attention by itself is permutation equivariant.
   B. Self-attention with position embeddings becomes permutation equivariant.
   C. Key, query and value transformations break permutation equivariance.

D. Multi-head attention breaks permutation equivariance.

7. Which of the following models is an autoregressive model:
   A. RealNVP
   B. WaveNet
   C. Generative Adversarial Networks
   D. Probabilistic PCA

8. Let us assume that x is a random variable that can take only binary values, i.e., $x \in \{0, 1\}$. In the VAE framework, what distribution we can use for the conditional likelihood $p(x|z)$:
   A. Bernoulli distribution
   B. Gaussian distribution
   C. Laplace distribution
   D. Poisson distribution

9. Let us assume an input $\mathbf{h} \in \mathbb{R}^D$ to a linear layer in a neural network with weights $\mathbf{W} \in \mathbb{R}^{D \times M}$, and a non-linearity $f(\cdot)$. How can we write a mathematical operation carried out by this linear layer using matrix calculus?
   A. $\mathbf{h}' = f(\mathbf{h}\mathbf{W})$
   B. $\mathbf{h}' = f(\mathbf{W}\mathbf{h})$
   C. $\mathbf{h}' = f(\mathbf{h}\mathbf{W}^\top)$
   D. $\mathbf{h}' = f(\mathbf{W}^\top\mathbf{h})$

10. An input to a Conv1D layer has length equal to 100. The kernel size is equal to 3, the stride is equal to 1. We want to obtain the output length equal to 100. What is the correct value of padding in this case?
    A. padding equal to 1
    B. padding equal to 2
    C. padding equal to 3
    D. padding equal to 4