

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

Prob #	1	2	3	4	Total
Points	25	25	25	25	

Time: 80 Minutes

Solution

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

$$L_i = \sum_{j \neq i} \max(0, y_j - y_i + \Delta)$$

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

$$H(p, q) = -\sum_x p(x) \log(q(x))$$

$$L_i = -\log\left(\frac{e^{y_i}}{\sum_j e^{y_j}}\right)$$

$$L_{KL} = -\frac{1}{2} \sum_{j=1}^2 \left[1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2 \right]$$

$$\text{Attention} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

$$\text{PE}(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right),$$

$$\text{PE}(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{10000^{2i/d_{\text{model}}}}\right)$$

CSE 4311 / CSE-5368 Neural Networks Spring 2026 Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

1. Consider a Transformer model with a maximum sequence length of **3** and an embedding dimension of **4**.

Given the sentence: **How are you**

Calculate the positional encoding for all the tokens.

Notes:

- position index starts from **0**
- Use the standard sinusoidal positional encoding. The equations are given on the back of the first page:
- Calculate all the numerical values to **3** digits of accuracy
- You **MUST SHOW** all your calculations for proper credit.

Given

- $d = 4$
- positions: 0,1,2
- $i = 0,1$

$$10000^{\frac{2i}{4}} = 10000^{i/2}$$

So:

- $i = 0 \Rightarrow 10000^0 = 1$
- $i = 1 \Rightarrow 10000^{0.5} = 100$

Position 0 ("How"): [0,1,0,1]

Position 1 ("are")

$$\left[\sin(1), \cos(1), \sin\left(\frac{1}{100}\right), \cos\left(\frac{1}{100}\right) \right] = [0.841, 0.540, 0.010, 0.9999]$$

Position 2 ("you")

$$\left[\sin(2), \cos(2), \sin\left(\frac{2}{100}\right), \cos\left(\frac{2}{100}\right) \right] = [0.909, -0.416, 0.020, 0.9998]$$

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

Problem 1 Continued

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

3. Consider a Transformer block with Embedding dimension = **4** ;
 Length of sentence = **2** ; Number of heads = **2** ; Number of blocks = **1**

Calculate the output for each head **with applied mask and softmax**

Note: Calculate all the numerical values to **3** digits of accuracy

- You **MUST SHOW** all your calculations including the attention scores, masked attentions, softmax attention scores, and attention outputs for **each head** for proper credit.

The input embedding matrix is: $X = \begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 2 \end{bmatrix}$

$$W_Q = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad W_K = \begin{bmatrix} 0 & 1 & -1 & -1 \\ -1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad W_V = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \\ -1 & 0 & 1 & -1 \end{bmatrix}$$

$$d_k = \frac{d_{model}}{h} = \frac{4}{2} = 2 \quad ; \quad Q = XW_Q, K = XW_K, V = XW_V$$

$$Q = \begin{bmatrix} 0 & -2 & 2 & -4 \\ 1 & 0 & -2 & 4 \end{bmatrix} ; \quad K = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix} ; \quad V = \begin{bmatrix} 4 & 2 & 0 & 2 \\ -3 & 0 & 1 & -3 \end{bmatrix}$$

Head 1

$$Q_1 = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \quad K_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad V_1 = \begin{bmatrix} 4 & 2 \\ -3 & 0 \end{bmatrix}$$

$$S_1 = \frac{Q_1 K_1^T}{\sqrt{2}}$$

$$\text{Scores: } S_1 = \begin{bmatrix} -1.414 & -1.414 \\ 0.707 & 0.000 \end{bmatrix}$$

$$\text{Masked Scores: } S_{1, \text{masked}} = \begin{bmatrix} -1.414 & -\infty \\ 0.707 & 0.000 \end{bmatrix}$$

$$\text{Softmax: } A_1 = \begin{bmatrix} 1.000 & 0.000 \\ 0.670 & 0.330 \end{bmatrix}$$

$$O_1 = A_1 V_1$$

$$\text{Output: } O_1 = \begin{bmatrix} 4.000 & 2.000 \\ 1.688 & 1.340 \end{bmatrix}$$

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

Problem 3 continued

Head 2:

$$Q_2 = \begin{bmatrix} 2 & -4 \\ -2 & 4 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 0 & 2 \\ 1 & -3 \end{bmatrix}$$

$$S_2 = \frac{Q_2 K_2^T}{\sqrt{2}}$$

$$\text{Scores: } S_2 = \begin{bmatrix} 1.414 & -1.414 \\ -1.414 & 1.414 \end{bmatrix}$$

$$\text{Masked scores: } S_{2,\text{masked}} = \begin{bmatrix} 1.414 & -\infty \\ -1.414 & 1.414 \end{bmatrix}$$

$$\text{Softmax: } A_2 = \begin{bmatrix} 1.000 & 0.000 \\ 0.056 & 0.944 \end{bmatrix}$$

$$O_2 = A_2 V_2$$

$$\text{Output: } O_2 = \begin{bmatrix} 0.000 & 2.000 \\ 0.944 & -2.721 \end{bmatrix}$$

$$\text{Final concatenated output: } O = \begin{bmatrix} 4.000 & 2.000 & 0.000 & 2.000 \\ 1.688 & 1.340 & 0.944 & -2.721 \end{bmatrix}$$

CSE 4311 / CSE-5368 Neural Networks Spring 2026Exam_02

Last Name	First Name	Student ID Number
Solution	«First_Name»	

4. Write a Python/NumPy function that computes the noisy image x_t at step t of the forward diffusion process. The forward diffusion equation is:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon$$

```
import numpy as np
def find_image_at_step(x0, t, betas):
    # Computes the noisy image x_t at diffusion step t.
    # x0 : np.ndarray Clean input image. Shape (H, W)
    # t : int Diffusion timestep.
    # betas : np.ndarray 1D-array of beta values for all timesteps.
    # Returns: xt : numpy.ndarray. Noisy image at timestep t.
    # Note: Noise should be standard normal Gaussian noise, same size
    as the image.
```

```
    noise=np.random.randn(*x0.shape)
```

```
    alphas=1.0-betas
```

```
    alpha_bars=np.cumprod(alphas)
```

```
    xt=np.sqrt(alpha_bars[t])*x0+sp.sqrt(1-alpha_bar[t])*noise
```

```
    return xt
```
