HW4

February 10, 2026

1 Homework 4

1.0.1 Wing Huang PID: A18894844

Date: 2026-01-27

1.0.2 Question 1: Setup and Weight Initialization

[1]:
```
import numpy as np
np.random.seed(42)
X = np.random.randn(128, 16)
print(X.shape)
```

```
(128, 16)
```

[2]:
```
W_Q = np.random.randn(16, 16)
W_K = np.random.randn(16, 16)
W_V = np.random.randn(16, 16)

print(W_Q.shape)
print(W_K.shape)
print(W_V.shape)
```

```
(16, 16)
(16, 16)
(16, 16)
```

1.0.3 Question 2: Linear Projections (Creating Q, K, V)

[3]:
```
Q = X @ W_Q
K = X @ W_K
V = X @ W_V

print(Q.shape)
print(K.shape)
print(V.shape)
```

```
(128, 16)
(128, 16)
(128, 16)
```

1.0.4 Question 3: The Attention Scores

[4]:
```
d_k = 16

scores = Q @ K.T
print(scores.shape)
```

```
(128, 128)
```

1.0.5 Question 4: The Softmax (Probability Distribution)

[5]:
```
def softmax(x):
    exp_x = np.exp(x)
    sum_exp = np.sum(exp_x, axis=1, keepdims=True)
    return exp_x / sum_exp
```

[7]:
```
attention_weights = softmax(scores)
print(attention_weights.shape)
print(np.sum(attention_weights, axis=1))
```

```
(128, 128)
[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1.]
```

1.0.6 Question 5: The Final Representation

[8]:
```
output = attention_weights @ V

print(output.shape)
```

```
(128, 16)
```

[ ]:
```

```