

**Problem 1**

0.0474

**Problem 2**

$W_3$

**Problem 3**

2. The softmax activation function is used in the output layer for multi-class classification problems to produce a probability distribution over multiple classes

**Problem 4**

3 .Leaky ReLU is a variant of ReLU that allows a small non-zero gradient for negative input values