1.

(1)Why do neural networks need activation functions?

1. Activation functions introduce **non-linearity** into the model. Without non-linear activation functions, a neural network, no matter how many layers it has, would simply behave as a linear model.
2. Real-world data often involves complex patterns that are non-linear in nature. Activation functions allow neural networks to **learn and model these intricate patterns effectively**.
3. Some activation functions, like ReLU, **introduce sparsity in the network** by outputting zero for negative input values. This sparsity can improve the efficiency of the network by focusing on a subset of active neurons and can also lead to better generalization by reducing overfitting.

(2) the influence of the value of the learning rate

- **Convergence Speed**: A high learning rate enables faster convergence, and a low learning rate makes training process slow.
- **Stability and Accuracy**: A high learning rate has risk of leading to divergence or oscillation around the minimum; A low learning rate ensures more stable and precise updates to the weights, while the training process may get stuck in local minima.
- **Training Dynamics**: A high learning rate can lead to erratic training dynamics, with significant fluctuations in the loss function; A low learning rate typically results in smoother training dynamics, with gradual and steady decreases in the loss function. This can be beneficial for achieving better generalization.

(3) What advantages does CNN have over fully connected DNN in image classification?

- **Capturing Local Features**: CNNs process small regions of the image through convolution layers, effectively recognizing local features like edges and textures, and learning spatial hierarchies of features.
- **Fewer Parameters**: Convolution operations use shared weights, significantly reducing the number of parameters, lowering computational complexity, and reducing the risk of overfitting.
- **Translation Invariance**: CNNs are robust to translations of objects in the image, allowing them to recognize objects regardless of their position.
- **Structural information & Hierarchical Feature Learning**: CNNs take into account more structural information of images, andlearn features from low-level to high-level, capturing complex patterns through their layered structure.
- **Higher Accuracy and Generalization**: CNNs typically achieve higher accuracy in image classification tasks and perform better on unseen data.

2

$N = 227$

$F = 11$

$Stride = 4$

$S_o = \frac{N-F}{Stride} + 1 = \frac{216}{4} + 1 = 55$

$Size = 55 \times 55 \times 96$

3

(1)

(a)

$$
\begin{aligned}
Val_{[1,1]} &= 1\times2+2\times0+3\times1+0\times0+1\times1+2\times2+3\times1+0\times0+1\times2=15\\
Val_{[1,2]} &= 2\times2+3\times0+0\times1+1\times0+2\times1+3\times2+0\times1+1\times0+2\times2=16\\
Val_{[2,1]} &= 0\times2+1\times0+2\times1+3\times0+0\times1+1\times2+2\times1+3\times0+0\times2=6\\
Val_{[2,2]} &= 1\times2+2\times0+3\times1+0\times0+1\times1+2\times2+3\times1+0\times0+1\times2=15
\end{aligned}
$$

The answer is

$$
\begin{bmatrix} 15 & 16 \\ 6 & 15 \end{bmatrix}
$$

(b)

$$
\begin{aligned}
Val_{[1,1]} &= 0\times2+0\times0+0\times1+0\times0+1\times1+2\times2+0\times1+0\times0+1\times2=7\\
Val_{[1,2]} &= 0\times2+0\times0+0\times1+1\times0+2\times1+3\times2+0\times1+1\times0+2\times2=12\\
Val_{[1,3]} &= 0\times2+0\times0+0\times1+2\times0+3\times1+0\times2+1\times1+2\times0+3\times2=10\\
Val_{[1,4]} &= 0\times2+0\times0+0\times1+3\times0+0\times1+0\times2+2\times1+3\times0+0\times2=2\\
Val_{[2,1]} &= 0\times2+1\times0+2\times1+0\times0+0\times1+1\times2+0\times1+3\times0+0\times2=4\\
Val_{[2,2]} &= 1\times2+2\times0+3\times1+0\times0+1\times1+2\times2+3\times1+0\times0+1\times2=15\\
Val_{[2,3]} &= 2\times2+3\times0+0\times1+1\times0+2\times1+3\times2+0\times1+1\times0+2\times2=16\\
Val_{[2,4]} &= 3\times2+0\times0+0\times1+2\times0+3\times1+0\times2+1\times1+2\times0+0\times2=10\\
Val_{[3,1]} &= 0\times2+0\times0+1\times1+0\times0+3\times1+0\times2+0\times1+2\times0+3\times2=10\\
Val_{[3,2]} &= 0\times2+1\times0+2\times1+3\times0+0\times1+1\times2+2\times1+3\times0+0\times2=6\\
Val_{[3,3]} &= 1\times2+2\times0+3\times1+0\times0+1\times1+2\times2+3\times1+0\times0+1\times2=15\\
Val_{[3,4]} &= 2\times2+3\times0+0\times1+1\times0+2\times1+0\times2+0\times1+1\times0+0\times2=6\\
Val_{[4,1]} &= 0\times2+3\times0+0\times1+0\times0+2\times1+3\times2+0\times1+0\times0+0\times2=8\\
Val_{[4,2]} &= 3\times2+0\times0+1\times1+2\times0+3\times1+0\times2+0\times1+0\times0+0\times2=10\\
Val_{[4,3]} &= 0\times2+1\times0+2\times1+3\times0+0\times1+1\times2+0\times1+0\times0+0\times2=4\\
Val_{[4,4]} &= 1\times2+2\times0+0\times1+0\times0+1\times1+0\times2+0\times1+0\times0+0\times2=3
\end{aligned}
$$

The answer is

$$\begin{bmatrix} 7 & 12 & 10 & 2 \\ 4 & 15 & 16 & 10 \\ 10 & 6 & 15 & 6 \\ 8 & 10 & 4 & 3 \end{bmatrix}$$

(2)

(a)

$$\begin{aligned} Val_{[1,1]} &= \max(1,4,5,8) = 8 \\ Val_{[1,2]} &= \max(2,1,3,4) = 4 \\ Val_{[2,1]} &= \max(7,6,1,3) = 7 \\ Val_{[2,2]} &= \max(4,5,1,2) = 5 \end{aligned}$$

The answer is

$$\begin{bmatrix} 8 & 4 \\ 7 & 6 \end{bmatrix}$$

(b)

$$\begin{aligned} Val_{[1,1]} &= \text{average}(1,4,5,8) = 4.5 \\ Val_{[1,2]} &= \text{average}(2,1,3,4) = 2.5 \\ Val_{[2,1]} &= \text{average}(7,6,1,3) = 4.25 \\ Val_{[2,2]} &= \text{average}(4,5,1,2) = 3 \end{aligned}$$

The answer is

$$\begin{bmatrix} 4.5 & 2.5 \\ 4.25 & 3.5 \end{bmatrix}$$