**Karan Vora kv2154**
**ECE-GY 7123 Introduction to Deep Learning**

**Problem 1):**

**Solution a):**

$$c=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

**Solution b):**

$$c=\begin{bmatrix}-1 & 0 & -1\\ -1 & 0 & -1\\ -1 & 0 & -1\end{bmatrix}$$

**Solution c):**

$$c=\begin{bmatrix}-1 & -1 & -1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix}$$

**Solution d):**

$$c=\begin{bmatrix}-1 & -1 & 0\\ -1 & 0 & -1\\ 0 & -1 & -1\end{bmatrix}$$

**Solution e):**

One example of an image operation that cannot be implemented using a 3x3 convolutional filter is a non-linear edge detector. A non-linear edge detector aims to highlight the edges in an image using non-linear operations. This can be achieved using techniques such as morphological operations or non-linear filtering, which involve operations such as dilation, erosion, and median filtering. These non-linear operations cannot be implemented using a 3x3 convolutional filter because a convolutional filter is a linear operator that operates on a local neighborhood of pixels in a fixed way. It applies the same linear transformation to each pixel in the neighborhood, regardless of its value or position. Therefore, a 3x3 convolutional filter cannot capture the complex non-linear relationships between pixels that are necessary for non-linear edge detection.

**Problem 2):**

**Solution a):**

The formula for $l_2$ loss is given by

$$L(w)=\frac{1}{2}\|y-Wx\|^2$$

So the new $l_2$ loss with $\lambda$ parameter for regularization is

$$\bar{L}(w)=L(w)+\lambda\|w\|_2^2$$

**Solution b):**

If the learning rate $\eta$, then the general gradient update rule is,

$$w_{t+1}=w_t-\eta\nabla\bar{L}(w)$$

$$w_{t+1}=w_t-\lambda\nabla L(w_t)-2\eta\lambda w_t$$

$$w_{t+1}=(1-2\eta\lambda)w_t-\eta\nabla L(w_t)$$

**Solution c):**

From the equation mentioned in Solution B, we can see that the updated $w$ consists of shrinking/decaying gradient by a factor of $(1-2\eta\lambda)$ and then updating in the direction of the gradient

**Solution d):**

Increasing $\lambda$ penalizes the $l_2$ norm of the weight vector, thus enforcing smaller weights on average. In order for the gradient to be stable, the constraining factor should be smaller than 1, i.e. $\eta<\frac{1}{2\lambda}$

**Problem 3):**

**Solution a):**

The definition of IOU for any two bounding boxes A and B is given by:

$$IOU(A,B)=\frac{|A\cap B|}{|A\cup B|}$$

Since the RHS is non-negative, the number has to be bigger than or equal to 0. Moreover, $A\cap B\subseteq A\cup B$ and hence the numerator has to be no bigger than the denominator. Therefore IOU is bounded between 0 and 1 (inclusive).

**Solution b):**

Consider two identical size square boxes A and B, both aligned at the same horizontal level. Fix B and then imagine sliding A from left to right. As A moves, the IOU will start from 0, increase until perfect overlap and then decrease until no overlap. The graph we will get is a step funtion i.e., it jumps from 0 to 1 when the boxes overlap and stays at 0 otherwise, the change in I)U will be discontinuous and will not have a well-defined derivative.