C-M-008: Machine Learning II '21-'22
HW 1 (Given December 8, 2021; Due December 11, 2021)

1) Consider the max pooling operation used in CNN's. For simplicity, let us look at a 1-D case and let the output of the previous (convolution) layer be 0 0 1 1 0 0 1 1 (usually the output of the convolution layer are real numbers but let us assume this in this hypothetical example). Assume you are max-pooling with a window width of 3 and a stride of 1. In one case you start max-pooling at location 1, and in the other case you start max-pooling at location 2. The second case in some sense willl occur when the input shifts by a position. Write the result of max-pooling in each of the two cases. If you see a difference what is the implication? Be very precise on the implication – you will not get any points if you are not precise (5+20 points)

input | 0 0 1 1 0 0 1 1 |

, Performing Max pooling on the input with windows width=1 and stride =1

Output = 1 1 1 1 1 1

As mentioned in question, starting from second position is same as shifting the input by a factor of 1.

New input = 0 1 1, 0, 0, 1, 1

Performing max Pooling, output becomes

= 1 1 1 1 1

if padding is used, we get same o/p

which is = 1 1 1 1 1

From the above example, the output remains the same, except for the change in length from 6 to 5 when padding is not used.
However, if we use a stride 2 and kernel size 2 in the above example, we can see the clear difference.

input = 0011001

output = 010

when pooling stash from 2nd position

output = 11

If we use a stride 1 in the above kernel ,

input = 0011001

output = 0111011

when pooling stash from 2nd position

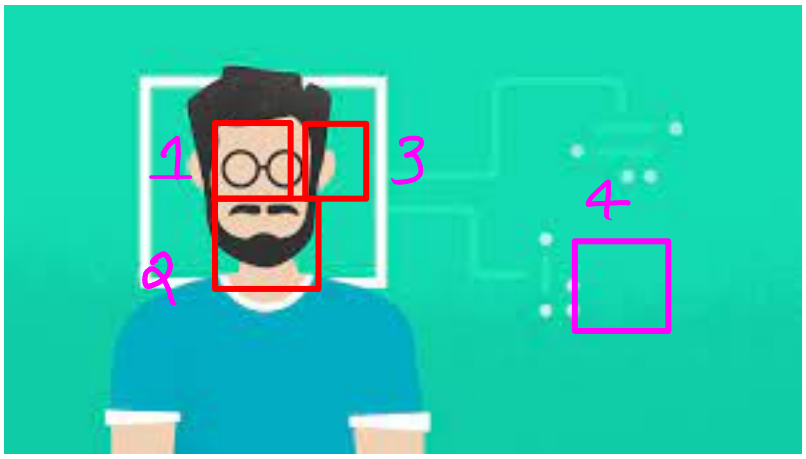
output = 11011

IMplications : When the stride is 1, the output is not impacted by the pooling operation. However, when stride is more than 1, we can see that the output varies with shift, ie is not shift invariant anymore.

On further reading, it was seen that the reason for that is that although convolutions are shift invariant, striding ignores the Nyquist sampling theorem and aliases which break the shift equivariance. This can be seen during Convolution and striding,

C-M-008: Machine Learning II '21-'22
HW 1 (Given December 8, 2021; Due December 11, 2021)

2) The convolution operation in CNN's is computed using pixels values in a neighborhood implying the presence of local features. Do you think this is a valid assumption? Explain your answer. Be precise. (10 points)



Yes, It is a valid assumption.

The pixels which are nearby when considered together helps in identifying features like ears, nose, vertical lines, horizontal lines etc. But pixels which are far from each other do not help in identifying any features as they do not have any relation with each other. Hence, it is valid to assume the presence of lcoal features in images.

When we look at the above image in the marked rectangles, we can see that the pixels which are in regions are correlated.
The pixels in region 1 helps in identifying feature, that it is eyes, pixels in region 2 helps in identifying ears etc.
But if we look at region 1 and region4, we cannot see any specific correaltion between the pixels as they do not help in identifying a feature.