DS-GA 1008 Homework 2 Question 1

Zian Jiang (zj444)

October 12, 2020

a

(1)

Convolution networks have 2 assumptions about their input data:

- **Locality:** it assumes that all signals are local. This assumptions make sense when dealing with data where each feature should be a function of a small part of space. For example, for images of faces, each features can be mouth, eyes or nose, which all only occupy a small portion of the image.
- **Stationarity:** That if one feature is useful to compute at some spatial position (x_1, y_1), then it should also be useful to compute at a different position (x_2, y_2). This allows us to share weights between different parts of the inputs.

(2)

Parameter sharing: because we assumed features are location invariant such that they can appear anywhere, this means that we can learn the same feature extractor for every location in an image, which results in massive savings in parameters as we don't have to relearn the same feature extractor for each location as would be the case for a fully connected neural network.

CNNs will fail when the input images have some specific centered structure, where completely different features should be learned on one side of the image than another. For example, inputs are faces that have been centered in the image. Different eye-specific or hair-specific features should be learned in different spatial locations.

b

$\frac{\partial z[k]}{\partial x[i]} = y[(i+k) \bmod n]$, since only one term out of the summation involves that particular $x[i]$.
$\frac{\partial z[k]}{\partial y[i]} = x[(i+n-k) \bmod n]$. This is because we can let $j = (i+k) \bmod n$, and rewrite i as $i = (j+n-k) \bmod n$.

C

$f(\cdot)$:

- **Dimensionality of output space:** $(1000,)$
- **Number of trainable parameters:** $1000 \times 100 = 10^5$
- **Computational complexity of forward pass:** $Wx^{(i)}$ involves 1000 dot products between 2 vectors of length 100, where each dot product involves 100 multiplications and 99 additions. Thus, in total $Wx^{(i)}$ involves $1000 \times (100 + 99) = 199,000$ operations. Pointwise non-linearity will be applied towards 1000 entries. So in total the foward pass involves $200,000$ operations.

$g(\cdot)$:

- **Dimensionality of output space:** $(\frac{100+2-3}{1} + 1, 10) = (100, 10)$
- **Number of trainable parameters:** $10 \times 3 = 30$
- **Computational complexity of forward pass:** This forward pass involves 10 feature maps, where each feature map involves 100 steps, and each step involves 3 multiplications and 2 additions, so 5 operations per step. Thus, before non-linearity, each feature map involves $5 \times 100 = 500$ operations. Then, non-linearity will be applied towards 100 numbers so in total for each feature map it involves $500 + 100 = 600$ operations. Across 10 feature maps, the total is $10 \times 600 = 6000$ operations.