AI5031: Machine learning, exercise 7

1 Math 1

Compute the derivatives $\frac{\partial a_j^{(l)}}{\partial a_k^{(l-1)}}$ for a ReLU layer, a layer that performs a transformation $g(x) = x^2$ and a softmax layer! Why does it make no sense to compute the derivative w.r.t. weight matrix and bias vector entries?

2 Math 2

Compute the derivative of an affine layer l activity $a_j^{(l)}$ w.r.t its bias vector entry $b_k^{(l)}$.

3 Parameters in fully-connected layers

Compute the number of parameters in a standard fully-connected (i.e., affine) layer l as a function of the sizes $Z^{(l)}$, $Z^{(l-1)}$ of layers l and $l-1$! Essentially, you need to compute the dimensions of the weight matrix and bias vector for this!

4 Parameters and dimensions in a convLayer

Assume a convLayer l, with the preceding convLayer having dimensions N, $W^{(l-1)} = H^{(l-1)} = 224, C^{(l-1)} = 3$. Assuming filter sizes $f_X^{(l)} = f_Y^{(l)} = 3$ and step sizes $\Delta_X^{(l)} = \Delta_Y^{(l)} = 1$:
a) Compute the dimensions of layer l with and without padding!
b) Compute the number of parameters in this layer if $C^{(l)}$ is chosen to be 64

5 VGG-16

Compute the number of parameters in VGG-16! Filter sizes are always 3x3, please refer to the lecture slides for the number of channels in each layer!

6 Implementing max-pooling

Write a Python function that takes a numpy array in NWHC format as a parameter, and performs max-pooling with a kernel size (same in both dimensions) that is a second parameter of that function. Test your function with arrays that are filled with ones, of the following dimensions, and print out the shape of the results:

a) 10,6,6,10, kernel size 3
b) 10,6,6,10, kernel size 2
c) 10,2,4,10, kernel size 2
d) 1,2,2,10, kernel size 2
Hint: please consider how the entries of the tensors are laid out in memory and solve this with a twofold reshape, followed by max() along a suitable axis. You can assume that the array dimensions are divisible by the kernel size!