Deep Learning: An Introduction for Applied Mathematicians (Part 2)

Gobind Puniani

MATH 190: Fall 2019

09/12/19

Convolutional Neural Networks

- Convolutional Neural Networks (CNNs) are a subset of ANNs, typically used in computer vision applications
- General ANNs would not be suitable for processing digital images due to sheer number of weights and biases
- CNNs apply linear kernels (filters) to portions of input data to create sparse matrices
  - Faster forward and backward passes
  - Filters can find structural changes in images (like edges)
- CNNs use 3-D tensors, where each pixel has 2 spatial coordinates and 1 RGB value

CNNs (Cont.)

- Called “convolutional” because the linear transformations can be written as convolutions
- 3-D tensors run through batches of input data and yield multi-dimensional output
- Pooling layers reduce dimensionality by assigning single numbers to small regions of pixels
  - Max pooling or average pooling

Training

- Overfitting is prevented by ceasing training once performance on validation data is no longer improving
- “Dropout” is also used to prevent overfitting by removing neurons to train smaller networks and then averaging them together in the end
- ReLU commonly used as activation function for image classification
- The final vector is run through the softmax operator to give probabilities for each category
- Results can be summarized in confusion matrices, in which off-diagonal entries indicate errors

Future Areas of Research

- Restarting the training process when significant mistakes are made
    - Rather than simply progressing with modified weights and biases
- Adversarial networks
    - Pitting a generative model against a discriminative model

Important Equations

- kth component of 1-D convolution: $y_k = \sum_{n=1}^{p} x_n g_{k-n}$
- ReLU function: $\sigma(x) = 0$, $x \geq 0$; $\sigma(x) = x$, $x > 0$
- Softmax operation: $(v^{\{i\}})_s \rightarrow \frac{e^{v_s^{\{i\}}}}{\sum_{j=1}^{K} e^{v_j^{\{i\}}}}$
- General cost function: Cost = $-\sum_{i=1}^{N} log\left(\frac{e^{v_{l_i}^{\{i\}}}}{\sum_{j=1}^{K} e^{v_j^{\{i\}}}}\right)$

Questions

- How were the five blocks constructed?
- How was the patch size determined?