Assignment 5 — Machine Learning

Tarush Goyal

December 15, 2020

1 Lab

1.3

1. **XOR** :
   (a) Average test accuracy for seeds 0 to 5 (both included) is 93.25%
   (b) Here I have a fully connected forward network with only 1 hidden layer of 5 neurons and relu activation. The output layer is of size 2 and has softmax activation.
   (c) number of epochs : 200
   (d) batch size : 100
   (e) learning rate : 0.0001
2. **Circle** :
   (a) Average test accuracy for seeds 0 to 5 (both included) is 91.15%
   (b) Here I have a fully connected forward network with only 1 hidden layer of 2 neurons and relu activation. The output layer is of size 2 and has softmax activation.
   (c) number of epochs : 200
   (d) batch size : 200
   (e) learning rate : 0.0001
3. **Mnist** :
   (a) Average test accuracy for seeds 0 to 4 (both included) is 94.27%
   (b) Here I have an input layer, a hidden layer of 256 neurons and relu activation and an output layer with 10 neurons and softmax activation.
   (c) number of epochs : 10
   (d) batch size : 50
   (e) learning rate : 0.0002
4. **cifar10** :
   (a) Test accuracy for seed = 335 is 37.4%
   (b) Here I have
      - a convolution layer with 4 filters of size (3,3) and stride 1
      - Average pooling of size (2,2) and stride 2

- a convolution layer with 4 filters of size (4,4) and stride 1
- Max pooling of size (2,2) and stride 2
- Flatten Layer
- Fully Connected layer with 10 output neurons and softmax activation

(c) number of epochs : 30

(d) batch size : 50

(e) learning rate : 0.0007

2 Theory

1.
   - The intermediate layers are mainly responsible for locating ”local parts” of the image. (the initial and final layers are responsible for edges and specifics patterns respectively)
   - In the first and third image, convolution layers would have identified various local parts like window, door, hood, roof, bumpers. In the second image, they would have identified handle, seat, fuel tank, clutch cover etc. Other parts like wheels, side mirrors and rims are common to both and would not contribute much and might have been rejected in initial layers. Sparse interactions and shared parameters are the properties of filters that are useful here.
   - Max and average pooling layers would work to ensure translation invariance among images. So that would let the model know that even image 3 is car though the local parts are shifted.

2.
   - We can use rectangular windows (subset of the image). We can iterate a fixed sized window over the image and apply the model in 1 (single label multi class classification) to each window. Wherever we have ”sufficient” confidence on our label, we can interpret that as a separate vehicle. eg. on an image of size 128 * 128, we can iterate 30 * 30 windows with a stride of 10 pixels. The drawback is that vehicles can be of different sizes and hence it is not easy to decide the window size and stride. Also we might need different window size within the same image which is very challenging.
   - another way of dealing with this would be to shift from single label classification to multi label classification. So we can replace softmax with a simple sigmoid activation in the output layer. That individual probabilities would be independent and we can set all probabilities ¿= 0.5 to be 1 and others 0. However a disadvantage is that the model has to handle all the ”local parts” together. So it can end up having low confidence on all the labels. Also it can mix various classes (eg. 2 parallel motorcycles could be predicted as car)

3. Here various vehicles would be partially visible inhibiting the model to see much ”local parts” and hence reducing confidence in many labels. So we can split the model into two parts, the first part detects various predefined ”local parts” and window in which they are present.Here we can use smaller windows (eg 10 x 10 in 128 x 128 images) than part 2 as we have to predict only parts and we can have more precision. Once this model has attained high accuracy for all windows, we use the results of this model as embeddings for training another convolution / fc model that predicts the vehicles in larger windows (eg. 30 x 30 in 128 x 128 images) This way we can account for not visible parts of vehicles and predict them in each window.
   Disadvantages would be :

- It is again difficult to choose window sizes for first and second parts of the model as the ideal size can vary image to image.
- Choosing local parts is also challenging. They have to be pre-labelled in training data which is expensive and time consuming.