Homework 5 Solutions

Ans 1.

(1) In character recognition task, flipping along the horizontal/vertical axis, rotation beyond a certain threshold, and random cropping are all unacceptable transformations as the label-preservation of the dataset is violated. E.g, 6 → rotated by 180* becomes 9, hence the dataset would become incorrect.

(2) Invariance such as the same object but in a different environment or context cannot be solved through data augmentation techniques. E.g.,

(a) Given an image of a landscape with a lake and mountains in summer and winter. Data augmentation might classify the winter landscape incorrectly as a glacier.

(b) Object rotated in 3D space and put in 2D, i.e., a side face profile wrt a front-face profile of the same person may get classified as 2 different objects.

Ans 2.

Input: h * w * c1

c1 >= h >= w >= 3

Convolution Type	Kernel Size	No. of Kernels	Stride	Output Size
1D	h * 2 * c1	c2	1	1 * (w-1) * c2
2D	3 * 3 * c1	c2	3	[floor(h-3)/3 + 1] * [floor(w-3)/3 + 1] * c2
3D	3 * 3 * 2	1	1	[floor(h-3) + 1] * [floor(w-3) + 1] * [floor(c1-2) + 1]

Ans 3.

Old architecture:

1. [227 * 227 * 3] INPUT
2. [55 * 55 * 96] CONV1: 96 11*11 filters at stride 4, pad 0
3. [27 * 27 * 96] MAX POOL1: 3*3 filters at stride 2
4. [27 * 27 * 96] NORM1: Normalization Layer
5. [27 * 27 * 256] CONV2: 256 5*5 filters at stride 1, pad 2
6. [13 * 13 * 256] MAX POOL2: 3*3 filters at stride 2
7. [13 * 13 * 256] NORM2: Normalization Layer

New architecture:

1. [227 * 227 * 3] INPUT

2. [55 * 55 * 96] CONV1: 96 11*11 filters at stride 4, pad 0
3. [55 * 55 * 96] 1*1 CONV: 96 1*1 filters at stride 1, pad 0
4. [55 * 55 * 96] 1*1 CONV: 96 1*1 filters at stride 1, pad 0

5. [27 * 27 * 96] MAX POOL1: 3*3 filters at stride 2
6. [27 * 27 * 96] NORM1: Normalization Layer

7. [27 * 27 * 256] CONV2: 256 5*5 filters at stride 1, pad 2
8. [27 * 27 * 256] 1*1 CONV: 96 1*1 filters at stride 1, pad 0
9. [27 * 27 * 256] 1*1 CONV: 96 1*1 filters at stride 1, pad 0

10. [13 * 13 * 256] MAX POOL2: 3*3 filters at stride 2
11. [13 * 13 * 256] NORM2: Normalization Layer

(1) Increase in the no. of parameters
= Parameters(3, 4) + Parameters(8, 9)
= 1*1 * 96*96 * 2 + 1*1 * 256*256 * 2
= 9216*2 + 65536*2
= 18432 + 131072
= 149,504

(2) Increase in memory (Assumption is that 1 parameter costs 4B for single precision)
= Memory of parameters + memory of output of hidden layers + memory of gradients (3, 4, 8, 9)
= (149,504 * 2 + 55*55*96*2 + 27*27*256*2) * 4B
= (299008 + 580800 + 373248) * 4B
= 1253056 * 4B
= 5,012,224 B
= 5.012 MB

Ans 4.

Condition	Bias	Variance
Adding weight decay	increases	decreases
Increasing the number of hidden units per layer	decreases	increases
Using dropout during training	increases	decreases
Adding more training data (drawn from the same distribution as before)	No change	decreases

- Regularization decreases variance and and increases bias

- Increasing model size reduces bias and increases variances
- No change in bias since the model learning capacity is same when training data is from the same distribution