Name - T Vijay
Roll no - CS17BTECH11040

1. Inlier ratio $w = 0.5$

Degrees of freedom in Homography = 8

$\therefore$ Minimum no. of sample required, $n = \lceil 8/2 \rceil = 4$

Let minimum no. of iterations required to achieve 95% probability of success = k

$$\therefore \quad 1 - (1 - w^n)^k > 0.95$$

$$0.05 > (1 - w^n)^k$$

$$k > \frac{\log(0.05)}{\log(1 - w^n)} = \frac{\log(0.05)}{\log(1 - (0.5)^4)}$$

$$k > 46.41$$

$\therefore$ Minimum iterations required is 47

2. $$\frac{\partial f}{\partial w^1_{ij}} = \left(\frac{\partial ((w^3)^T h^{(2)})}{\partial w^1_{ij}} \right)$$

$$= (w^3)^T \frac{\partial h^{(2)}}{\partial w^1_{ij}}$$

We know $h^2 = \sigma(w^2 h^1)$

and $\frac{d}{dx}\sigma(x) = \sigma(x)(1-\sigma(x))$

$$\frac{\partial f}{\partial w^1_{ij}} = (w^3)^T \left(\sigma(w^2 h^{(1)})(1-\sigma(w^2 h^1)) \right) \left(\frac{\partial w^2 h_1}{\partial w^1_{ij}} \right)$$

$$= (w^3)^T \left(\sigma(w^2 h^{(1)})(1-\sigma(w^2 h^1)\right) w^2 \left(\sigma(w^1 x)(1-\sigma(w^1 x)) \frac{\partial w^1 x}{\partial w^1_{ij}} \right)$$

$$(\because h^1 = \sigma(w^{(1)} x))$$

$$\frac{\partial f}{\partial w^1_{ij}} = (w^3)^T (\sigma(w^2 h^1)(1-\sigma(w^2 h^1)) w^2 \sigma(w^1 x)(1-\sigma(w^1 x)) x$$

Scanned with CamScanner

Q3

$$\Delta^{(2)} := \Delta^{(2)} + \delta^{(3)} (a^{(2)})^T$$

Q4

Total weights $= M \times d + C \times M = M(c+d)$

Total biases $= M + C$

Total independent ~~der~~ derivate $= M + C$

(we need to compute $\delta^{(2)}$ & $\delta^{(3)}$)

Scanned with CamScanner

5. Since ϵ_n is $\mathcal{N}(0, \Sigma)$

$\Rightarrow y_n$ is $\mathcal{N}(f(x_n, w), \Sigma)$

$(\because y_n = f(x_n, w) + \epsilon_n)$

$\therefore P(y_n = y / x_n)$ = Probability distribution of Gaussian distribution

$$\therefore P(y_n = y/x_n) \propto \exp(-(y - f(x_n; w))^T \Sigma^{-1} (y - f(x_n, w)))$$

$$P(Y/X) = \prod_i^N P(y_i/x_i)$$

$$\propto \exp\left(-\sum_i^N (y - f(x_i, w))^T \Sigma^{-1} (y - f(x_i, w))\right)$$

And we know minimizing the sum of squared error is equivalent to MLE

In MLE, we try to ~~maximize~~

$$\max \sum_i^N \log P(Y_i / x_i, w)$$

$$\therefore L(w) = -\log P(Y/X)$$

$$= \sum_i^N (y_i - f(x_i, w))^T \underbrace{\Sigma^{-1}}_{\sigma^2 I} (y_i - f(x_i, w))$$

$\therefore$ This is equivalent to sum of squared error when $\Sigma = \sigma^2 I$

Scanned with CamScanner

6. a) The problem caused by scale symmetry is vanishing & exploding gradients.

Proof:

Let w_1, w_2 be weights of two layers

$$L(D, w_1, w_2) = L(D, \gamma w_1, w_2/\gamma)$$

(scale symmetry)

Let $\gamma w_1 = P_1$ & $w_2/\gamma = P_2$

$$\therefore L(D, w_1, w_2) = L(D, P_1, P_2)$$

$$\frac{\partial L(D, w_1, w_2)}{\partial w_1} = \frac{\partial L(D, P_1, P_2)}{\partial w_1}$$

$$= \frac{\partial L}{\partial P_1}(D, P_1, P_2) \cdot \gamma$$

$\therefore$ Exploding gradient can be seen at layer 1

and similarly

$$\frac{\partial L(D, w_1, w_2)}{\partial w_2} = \frac{\partial L}{\partial P_2}(D, P_1, P_2)\frac{1}{\gamma}$$

$\therefore$ Vanishing gradient

$\therefore$ Poor training and learning.

Scanned with CamScanner

b) Value symmetry: when we initial all weights and biases with same value, the model doesn't learn.

This because the backpropagated error is proportional to the values of the weights.

Scanned with CamScanner