Name - Shivashish Suman

Roll No. - CS17BTECH11037

1. Let w denote (number of inliers in data / number of points).

$\therefore$ w is the probability of choosing an inlier.

We assume that n points are needed for estimating a model.

Here $w = 0.5$ (given)

Probability that algorithm never selects a set of n points which all are inliers $= 1 - 0.95 = 0.05$ (given)

Also it is equal to $(1-w^n)^K$ where k is equal to the number of iterations. This is because w^n is probability that all n points are inliers and $1-w^n$ is prob. that atleast one point is ~~an~~ outlier and $(1-w^n)^K$ is the probability that atleast one outlier is present in all iterations.

$$\therefore \quad (1-(0.5)^n)^K = 0.05$$

$$\Rightarrow \quad K \log_2(1-0.5^n) = \log_2 0.05$$

$$\Rightarrow \quad K = \frac{\log_2(0.05)}{\log_2(1-0.5^n)} = \text{Number of iteration}$$

In case of homography, $n = 4$ $\quad \therefore K \approx 47$.

2. $$\frac{\partial f}{\partial w_{i,j}^{1}} = \frac{\partial f}{\partial h_j^{1}} \cdot \frac{\partial h_j^{1}}{\partial w_{i,j}^{1}} = \frac{\partial f}{\partial h_j^{1}} \sigma'\left(\sum_i x_i w_{ij}^{1}\right) \cdot x_i$$

$$\left\{\sigma'(y) = \sigma(y)\left(1-\sigma(y)\right)\right\}$$

$$= \frac{\partial f}{\partial h_j^{1}} \sigma'(z_j^{1}) \cdot x_i$$

$$\left\{ \sum_i x_i w_{ij}^{1} = z_j^{1} \right\}$$

$$\frac{\partial f}{\partial h_j^{1}} = \frac{\partial f}{\partial h_1^{2}} \cdot \frac{\partial h_1^{2}}{\partial h_j^{1}} + \frac{\partial f}{\partial h_2^{2}} \cdot \frac{\partial h_2^{2}}{\partial h_j^{1}}$$

$$= w_{1}^{3} \frac{\partial h_1^{2}}{\partial h_j^{1}} + \cancel{\frac{\partial f}{\partial h_2^{2}}} w_2^{3} \frac{\partial h_2^{2}}{\partial h_j^{1}}$$

$$= w_1^{3} \sigma'(z_1^{2}) w_{j1}^{(2)} + w_2^{3} \sigma'(z_2^{2}) w_{j2}^{2}$$

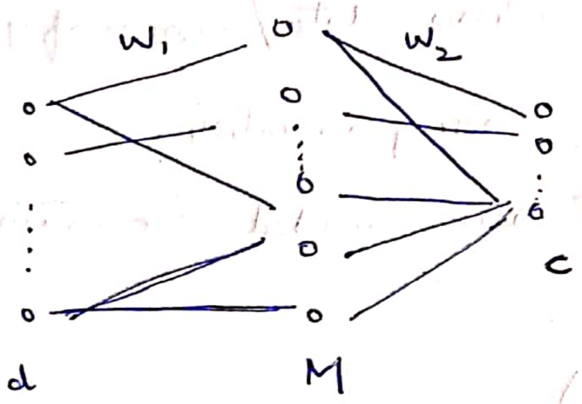
$$\left\{ \because \frac{\partial f}{\partial h_1^{2}} = w_1^{3}, \quad \frac{\partial f}{\partial h_2^{2}} = w_2^{3} \right.$$

$$\therefore \frac{\partial f}{\partial w_{i,j}^{1}} = \sigma'(z_j^{1}) \cdot x_i \left[w_1^{3} \sigma'(z_1^{2}) w_{j1}^{2} + w_2^{3} \sigma'(z_2^{2}) w_{j2}^{2} \right]$$

$$\left\{ \text{Here } z_1^{2} = \sum_{j=1}^{2} w_{j1}^{2} h_j^{1}, \quad z_2^{2} = \sum_{j=1}^{2} w_{j2}^{2} h_j^{1} \right\}$$

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

4.



So in W_1, there will be M biases and $M*d$ weights

Similarly in W_2, there will be c biases and $M*c$ weights

$\therefore$ Total weights and biases $= M + M*d + c + M*c$

No. of independent derivative is equal to number of $\delta^{(i)}$ values for $i=1,2$ Here $\delta^{(1)}$ contains M entries and $\delta^{(2)}$ contains c entries so in total $M+c$.

$\therefore$ No. of independent derivatives $= M+c$

5. Likelihood of an iid sequence $X = \{x_1 .. x_n\}$ is given by

$$L(w) = f(X; w) = \prod_{i=1}^{n} \left(f(x_i; w) + \epsilon_i \right)$$

Error is given by $\left(y - f(x_i; w)\right)^T \Sigma^{-1} \left(y - f(x; w)\right)$

6 (a) Scaling up weights in one layers and scaling down in other layer might lead to giving more weightage to less prominent feature and hence although the loss function may remain same for training set, it may perform poorly on test data set.

(b) one symmetry is weight initialization with all weights and biases equal to 1. In this, all weights ~~are~~ might get updated in a similar fashion and hence they will continue to remain similar in value and neural network will fail to learn.