Solution-1

It is given that we wish to have 95% chance of computing correct homography.

⊙ so, Probablity of failure = $1-0.95 = 0.05$

⊙ number of feature pair required to calculate homography = 4
$n = 4$

⊙ Probablity of failure = Probablity of never selecting n points of inliers.

Probablity of never selecting n points of inlier = $(1-w^n)^K$

⊙ $(1-w^n)^K = 0.05$

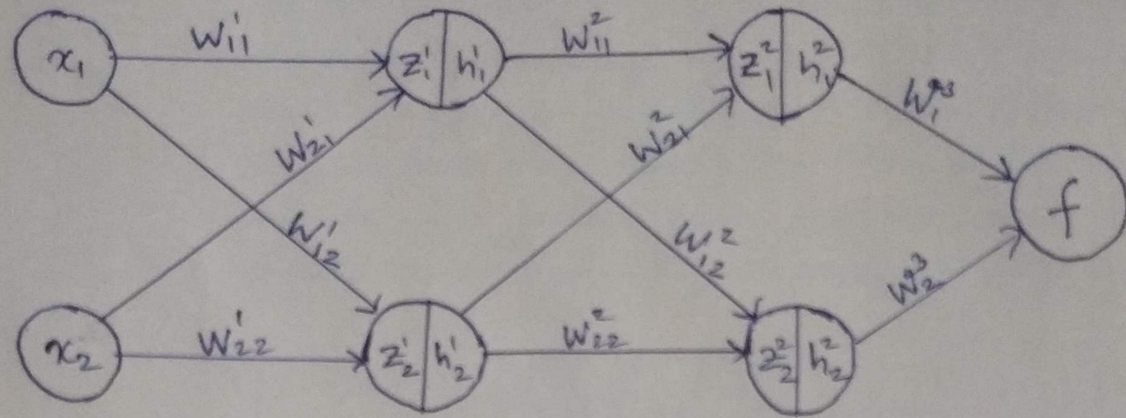
⊙ $(1-(0.5)^4)^K = 0.05$

⊙ $(0.9375)^K = 0.05$

⊙ $$\boxed{K = \frac{\log_e 0.05}{\log_e 0.9375} = 46.41 \approx 47}$$

so, the number of itterations required are 47 for 0.95 probablity of selecting 4 feature points as inliers.

Solution-2



(1) $$\frac{\partial f}{\partial w^1_{11}} = \left(\frac{\partial f}{\partial h^2_1}\cdot\frac{\partial h^2_1}{\partial z^2_1}\cdot\frac{\partial z^2_1}{\partial h^1_1} + \frac{\partial f}{\partial h^2_2}\cdot\frac{\partial h^2_2}{\partial z^2_2}\cdot\frac{\partial z^2_2}{\partial h^1_1}\right)\cdot\frac{\partial h^1_1}{\partial z^1_1}\cdot\frac{\partial z^1_1}{\partial w^1_{11}}$$

$$\boxed{\frac{\partial f}{\partial w^1_{11}} = \left(w^3_1\cdot\sigma'(z^2_1)\cdot w^2_{11} + w^3_2\cdot\sigma'(z^2_2)\cdot w^2_{12}\right)\sigma'(z^1_1)\, x_1} \rightarrow (1)$$

(2) $$\frac{\partial f}{\partial w^1_{21}} = \left(\frac{\partial f}{\partial h^2_1}\cdot\frac{\partial h^2_1}{\partial z^2_1}\cdot\frac{\partial z_1}{\partial h^1_1} + \frac{\partial f}{\partial h^2_2}\cdot\frac{\partial h^2_2}{\partial z^2_2}\cdot\frac{\partial z^2_2}{\partial h^1_1}\right)\frac{\partial h^1_1}{\partial z^1_1}\cdot\frac{\partial z^1_1}{\partial w^1_{21}}$$

$$\boxed{\frac{\partial f}{\partial w^1_{21}} = \left(w^3_1\cdot\sigma'(z^2_1)\cdot w^2_{11} + w^3_2\cdot\sigma'(z^2_2)\cdot w^2_{12}\right)\cdot\sigma'(z^1_1)\cdot x_2} \rightarrow (2)$$

(3) $$\frac{\partial f}{\partial w^1_{22}} = \left(\frac{\partial f}{\partial h^2_2}\cdot\frac{\partial h^2_2}{\partial z^2_2}\cdot\frac{\partial z^2_2}{\partial h^1_2} + \frac{\partial f}{\partial h^2_1}\cdot\frac{\partial h^2_1}{\partial z^2_1}\cdot\frac{\partial z^2_1}{\partial h^1_2}\right)\cdot\frac{\partial h^1_2}{\partial z^1_2}\cdot\frac{\partial z^1_2}{\partial w^1_{22}}$$

$$\boxed{\frac{\partial f}{\partial w^1_{22}} = \left(w^3_2\cdot\sigma'(z^2_2)\cdot w^2_{22} + w^3_1\cdot\sigma'(z^2_1)\cdot w^2_{21}\right)\cdot\sigma'(z^1_2)\cdot x_2} \rightarrow (3)$$

4) $$\frac{\partial f}{\partial W_{12}^{1}} = \left(\frac{\partial f}{\partial h_2^2}\cdot\frac{\partial h_2^2}{\partial z_2^2}\cdot\frac{\partial z_2^2}{\partial h_2^1} + \frac{\partial f}{\partial h_1^2}\cdot\frac{\partial h_1^2}{\partial z_1^2}\cdot\frac{\partial z_1^2}{\partial h_2^1}\right)\frac{\partial h_2^1}{\partial z_2^1}\cdot\frac{\partial z_2^1}{\partial W_{12}^1}$$

$$\boxed{\frac{\partial f}{\partial W_{12}^1} = \left(W_2^3\cdot\sigma'(z_2^2)\cdot W_{22}^2 + W_1^3\cdot\sigma'(z_1^2)\cdot W_{21}^2\right)\cdot\sigma'(z_2^1)\cdot x_1} \rightarrow (4)$$

So, in General form $\frac{\partial f}{\partial W_{ij}^1}$ can be written as :-

$$\boxed{\frac{\partial f}{\partial W_{ij}^1} = \left(W_2^3\cdot\sigma'(z_2^2)\cdot W_{j2}^2 + W_1^3\cdot\sigma'(z_1^2)\cdot W_{j1}^2\right)\cdot\sigma'(z_j^1)\cdot x_i} \rightarrow (5)$$

Solution - 3

update equation $\Delta_{ij}^{(2)} = \Delta_{ij}^{(2)} + \delta^{(3)}_i (a^{(2)})_j$

can be written in vector form as :-

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

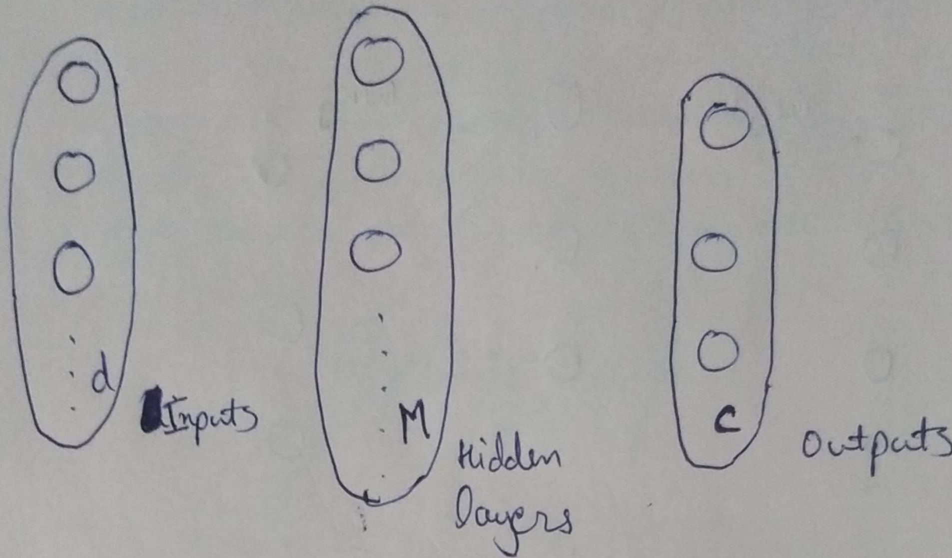
where $\delta^{(3)} \cdot (a^{(2)})^T = \left[\delta_i^{(3)} \cdot (a^{(2)})_j \right]$

is a matrix where every ij ~~element~~ index is the element which we wanted

$$\delta^{(3)} \cdot (a^{(2)})^T = \begin{bmatrix} \delta_1^{(3)} \cdot (a^{(2)})_1 & \delta_1^{(3)} \cdot (a^{(2)})_2 & \cdots \\ \delta_2^{(3)} (a^{(2)})_1 & \delta_2^{(3)} \cdot (a^{(2)})_2 & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

nodes in i layer
× nodes in j layer

Solution-4



Each Input will be connected with each node in Hidden layers, so Number of weights $= d \times M$

Each node of ~~output~~ Hidden layer will be connected to node of output layer, so no. of weights $= M \times C$

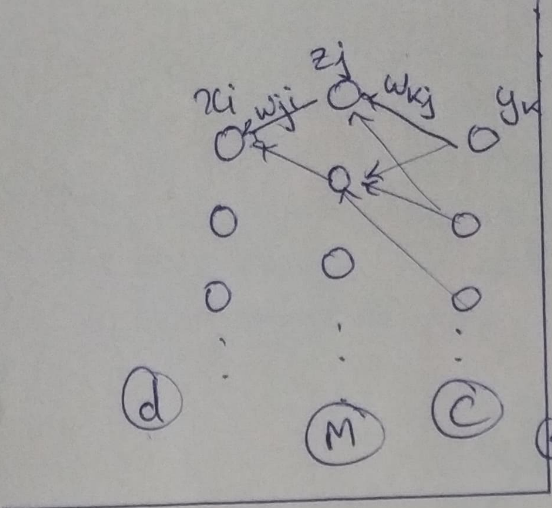
(i)

Total no. of weights in network $= d \times M + M \times C$
$= (d+c)M$

(ii)

Total number of bias will be $=$ Nodes in Hidden layer + Nodes in output layer
$= (M+C)$

Sol-4 Continued...



Considering derivative of error function for weights of one input examples only.

- for node (z_j) connecting (x_i) input. weight is w_{ji}

- error function derivative w.r.t w_{ji} $\left(\frac{\partial E}{\partial w_{ji}}\right) = \sum_{k=1 \text{ to } c} \frac{\partial E}{\partial w_{kj}}$

$$= \sum_{k=1 \text{ to } c} \delta_k . z_j$$

- derivative ~~of all~~ w.r.t all weights that connect to input x_i $= \sum_{j=1 \text{ to } M} \left(\delta_j . x_i \left(\sum_{k=1 \text{ to } c} \delta_k . z_j \right) \right)$

- total number of derivatives of activation function $= M$.

So to ~~total~~ number of independent derivates

$$= M \times \underline{(M \times C)}$$

$\downarrow$ derivatives for activation function used

Solution - 5

- Given is the target data of the form

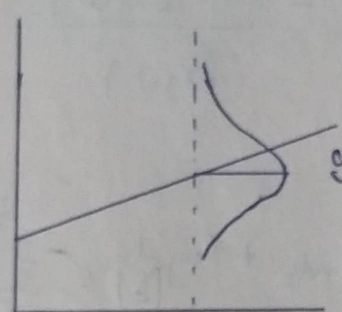
$$y_n = f(x_n; w) + \varepsilon_n$$

where ε_n is error with Gaussian distribution

$$\varepsilon_n \in N(0, \Sigma)$$

- y_n can also be represented as Gaussian distribution such as mean is $f(x_n; w)$ & variance is Σ.

$$y_n \in N(f(x_n; w); \Sigma)$$



y_i is normally distributed with mean as $f(x_n; w)$

- Likelyhood of observing y_i can be written as $= P(y_i \mid f(w, x_i))$
- Likelyhood of observing all such target points $= \prod_{i=1}^{n} P(y_i \mid f(w, x_i))$

$$\prod_{i=1}^{n} P(y_i \mid f(w, x_i)) = \prod N(y_i \mid f(x_i; w); \Sigma)$$

$$L(w) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}} \times \frac{1}{|\Sigma|} e^{-\frac{1}{2}\left[(y_i - f(x_i; w))^T \Sigma^{-1} (y_i - f(x_i; w))\right]}$$

$$\boxed{L(w) = \frac{1}{(2\pi)^{nd/2}} \times \frac{1}{|\Sigma|} e^{-\frac{1}{2}\left[\sum_{i=1}^{m} (y_i - f(x_i; w))^T \Sigma^{-1} (y_i - f(x_i; w))\right]}}$$

This is our likelyhood function.

- Taking log of likely hood function.

$$ll(w) = \log \frac{1}{(2\pi)^{nd/2}} + \log \frac{1}{|\Sigma|^{n/2}} - \frac{1}{2} \sum_{i=1}^{n} \left[(y_i - f(x_i;w))^T \Sigma^{-1} (y_i - f(x_i;w)) \right]$$

- We know that maximising log likely hood ($ll(w)$) function is equivalent to minimising sum of squared error.

$$\frac{\partial(E(w))}{\partial(w)} = -\frac{\partial\, ll(w)}{\partial(w)}$$

$$\frac{\partial(E(w))}{\partial(w)} = -\frac{\partial}{\partial(w)} \left[\log \frac{1}{(2\pi)^{nd/2}} + \log \frac{1}{|\Sigma|^{n/2}} - \frac{1}{2} \sum_{i=1}^{n} \left[(y_i - f(x_i;w))^T \Sigma^{-1} (y_i - f(x_i;w)) \right] \right]$$

$$\frac{\partial(E(w))}{\partial(w)} = -\left[0 + 0 - \frac{1}{2} \frac{\partial}{\partial(w)} \sum_{i=1}^{n} \left[(y_i - f(x_i;w))^T \Sigma^{-1} (y_i - f(x_i;w)) \right] \right]$$

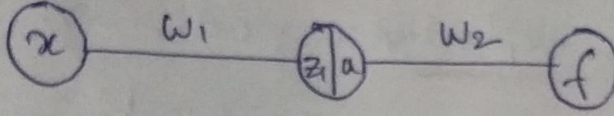
integration both sides

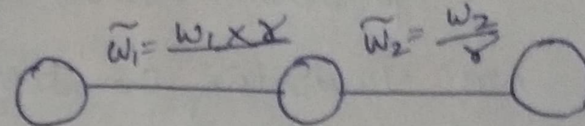
$$\int_w \frac{\partial E(w)}{\partial(w)} = \int_w \frac{1}{2} \frac{\partial}{\partial(w)} \sum_{i=1}^{n} \left[(y_i - f(x_i;w))^T \Sigma^{-1} (y_i - f(x_i;w)) \right]$$

$$\boxed{E(w) = \frac{1}{2} \sum \left[(y_i - f(x_i;w))^T \Sigma^{-1} (y_i - f(x_i;w)) \right]}$$

this is the required error function

6(a) The problem caused by Scale Symmetry is of Vanishing Gradients & exploding gradients. We can see this with simple networks.





in forward pass.

- $z_1 = w_1 x$
- $a_1 = \mathrm{ReLU}(w_1 x)$
- $f = a_1 \times w_2$

in forward pass

- $\tilde{z}_1 = \gamma w_1 x = \gamma z_1$
- $\tilde{a}_1 = \mathrm{ReLU}(\gamma w_1 x) = \gamma a_1$
- $\tilde{f} = \tilde{a}_1 \times \tilde{w}_2 = \gamma a_1 \times \frac{w_2}{\gamma} = f$

we see that output f is same in both cases. therefore Loss function (L) will also be same for both

gradients

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial f} \cdot \frac{\partial f}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$

gradients

$$\frac{\partial L}{\partial \tilde{w}_1} = \frac{\partial L}{\partial \tilde{f}} \cdot \frac{\partial \tilde{f}}{\partial \tilde{a}_1} \cdot \frac{\partial \tilde{a}_1}{\partial \tilde{z}_1} \cdot \frac{\partial \tilde{z}_1}{\partial \tilde{w}_1}$$

$$\frac{\partial L}{\partial \tilde{w}_1} = \frac{\partial L}{\partial f} \cdot \frac{\partial f}{\partial (\gamma a_1)} \cdot \frac{\partial (\gamma a_1)}{\partial (\gamma z_1)} \cdot \frac{\partial (\gamma z_1)}{\partial (\gamma w_1)}$$

$$\frac{\partial L}{\partial \tilde{w}_1} = \frac{\partial L}{\partial f} \cdot \frac{\partial f}{\gamma (\partial a_1)} \cdot \frac{\cancel{(\gamma)} \partial a_1}{\cancel{(\gamma)} \partial z_1} \cdot \frac{\cancel{(\gamma)} \partial z_1}{\cancel{(\gamma)} \partial w_1}$$

$$\boxed{\frac{\partial L}{\partial \tilde{w}_1} = \frac{1}{\gamma} \frac{\partial L}{\partial w_1}}$$

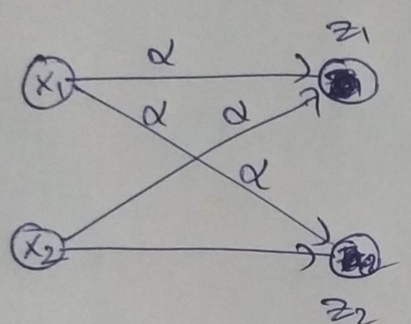
Hence we find that gradients can vanish for large value of γ. and gradients can explode for small values of γ.

6 (b)

Another kind of Symmetry can be weight symmetry with in adjacent layers.

this is a problem in many case as the network is not able to learn effectively.

this can be backed up with a simple example



$$\boxed{\begin{aligned} z_1 &= \alpha x_1 + \alpha x_2 \\ z_2 &= \alpha x_1 + \alpha x_2 \end{aligned}}$$

↓

this implies that both the layers will have similar effect on the cost function.

And we know that this will now give same gradients during back propogation. Hence

* Both the neuron in hidden layer will learn Similar things which is a problem for us.