1)

Assigment 2 : Deep Learning : AI2100

RAJ PATIL : CS18BTECH11039

1.) given that 50% of the initial matches are correct i.e. inliers, outlier rate is 0.5.

Now we need to randomly sample $\underline{4}$ correspondences for each iteration of RANSAC.

let N be the required number of iterations in order to generate a 95% chance to generate the correct homography.

Some definitions

$p \triangleq 0.95$, $m \triangleq 4$

↳ desired chances

$u \rightarrow$ probability of inlier = 0.5

$v \rightarrow$ probability of outlier = 0.5

∴ for a good match finding N iter s.t. atleast 1 iteration has no outliers

$$1-p = (1-u^m)^N$$

$$\therefore N = \frac{\log(1-p)}{\log(1-u^m)} = \frac{\log(0.05)}{\log(1-0.5^4)}$$

$$= 46.4$$

∴ $\boxed{47 \text{ iterations}}$

2) let δ_i^l represent error term corresponding to l^{th} layer's i^{th} node

then

$$\nabla_{w_{ij}^1} f = (\delta_i^2)(x_j) \quad \cdots \quad (1)$$

where $$(\delta_i^2) = \left(\sum_{j=1}^{2} w_{ij}^2 \, \delta_j^3\right)(h_i^1)'$$

where $$(h_i^1)' = (h_i^1)(1 - h_i^1) \quad \text{(sigmoid)}$$

& $$(\delta_j^3) = (w_j^3 f)(h_j^2)'$$

where $$(h_j^2)' = (h_j^2)(1 - h_j^2) \quad \text{(sigmoid) again}$$

$\therefore$ substituting all back in (1)

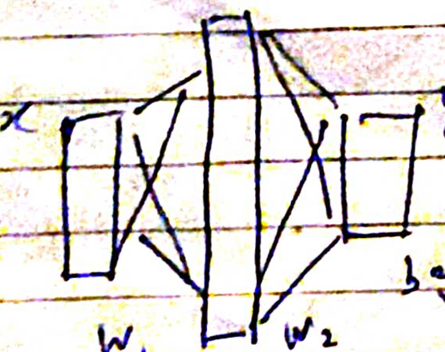
$$\boxed{\frac{\partial f}{\partial w_{ij}^1} = \left(\sum_{j=1}^{2} w_{ij}^2 (w_j^3 f)(h_j^2)(1 - h_j^2)\right)(h_i^1)(1 - h_i^1)(x_j)}$$

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

Outer product of the two vectors

gives matrix of size equal to that of Δ^2 & ij^{th} element as desired.

4) $a' \in \mathbb{R}^M$ $y \in \mathbb{R}^C$ $x \in \mathbb{R}^d$



$b_2 \rightarrow + C$

$W_2 \rightarrow C \times M$

$W_1 \rightarrow M \times d$, $b_1 \rightarrow + M$

for W_1 for W_2

$Md + CM$

$+ M + C$

for b_1 for b_2 (biases)

$\boxed{Md + CM + M + C}$

~~202~~

Page No.
Date

5.) $y_n - f(x_n; w) \sim \mathcal{N}(0_n, \Sigma)$

given a pair of ~~points~~ features & output x_n, y_n for a dataset

The Likelihood of observing that dataset could be given by

$$\cancel{\mathcal{L}((x, y)) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{-(y - f(x_n, w) - 0)}}$$

$$\cancel{\frac{1}{2}}$$

$$\mathcal{L}((x_n, y_n)) = \frac{\exp\left(-\frac{1}{2}(y_n - f(x_n, w) - 0_n)^T \Sigma^{-1} (y_n - f(x_n, w) - 0_n)\right)}{\sqrt{(2\pi)^n |\Sigma|}}$$

$(\mu_n = 0_n)$

$$= \frac{\exp\left(-\frac{1}{2}(y_n - f(x_n, w))^T \Sigma^{-1} (y_n - f(x_n, w))\right)}{\sqrt{(2\pi)^n |\Sigma|}}$$

Maximizing ~~Likelihood~~ log Likelihood taking log

~~log~~ (for ease) → log is monotonic

maximum doesn't change.

$$\ln(L) = c - \frac{1}{2}(y_n - f(x_n,w))^T \Sigma^{-1} (y_n - f(x_n,w))$$

$\therefore$ The corresponding loss function will be NLL (negative log likelihood)

given by

$$\boxed{E = \frac{1}{2}(y_n - f(x_n,w))^T \Sigma^{-1} (y_n - f(x_n,w))}$$

for $\Sigma = \sigma^2 I$ we have

$$E = \frac{1}{2\sigma^2} \| y_n - f(x_n,w) \|_2^2$$

= sum of least squares.

6-> a)

Given two layers related ~~as~~ by a scale γ, the following issue could arise

~~if γ is too large~~

if γ is such that one layers is scaled up a lot & the other layer is scaled down correspondingly, one can observe the following

- In the layer being scaled up, we lose precision for representing the weights (~~as~~ 32 bit or 64 bit floats) as ~~at~~ normal standards like IEEE 754 are such that numbers space out as they get larger :- this might not provide the granularity needed for learning efficiently
- In the layer being scaled down, some weights might go down to zero as computers again cannot represent all ~~[illegible]~~ real numbers.

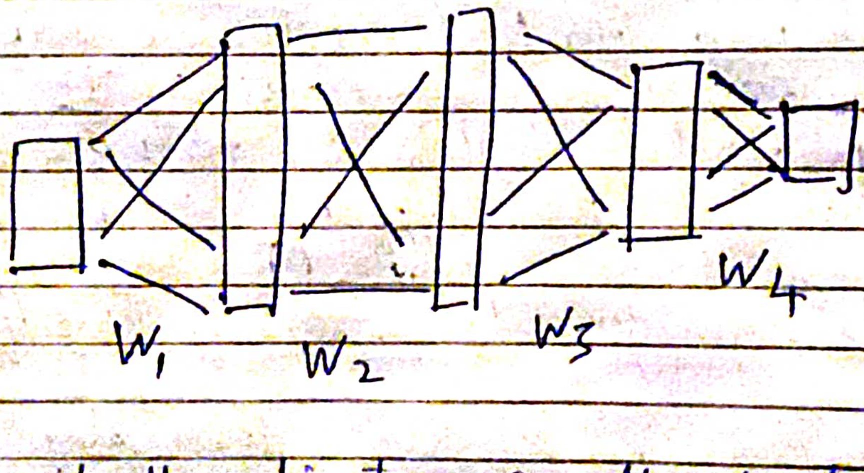
Another issue that could arise is that due to multiple models in the hypothesis space performing the same way will result in difficulties for convergence.

Page No.
Date / /

6) b) Consider a neural network with only linear activations then one can observe, what one can call, sign-symmetry

if ~~two~~ consec the signs of an even no. of layers in the network are flipped, ~~the output will still be the~~ same.

~~eg~~ eg :-



if ~~all~~ activations are affine functions then we have a total of atleast

$$2\left(\binom{4}{4}+\binom{4}{2}\right)=2\cdot\cancel{7}=14$$ hypotheses that behave in the same way.

$\binom{4}{4}$ → choose 4 to flip

$\binom{4}{2}$ → choose 2 to flip