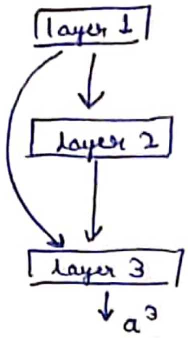
Name - Shivashish Suman

Roll No. - CS17BTECH11037

1. Consider Resnet Model :



Let $W^{2,3}$ be weight matrix for layer 2 to layer 3.

Let $W^{1,3}$ be weight matrix for connecting layer 1 to 3.

∴ Forward propogation is given by -

$$a^3 = f(z^3 + w^{1,3} \cdot a^1)$$

where $a^{(i)}$ is activations of neurons in layer i.

f is activation function of layer 3.

$$z^3 = W^{2,3} a^2 + b^3$$

Now coming to back propogation :

$$\Delta_{w^{2,3}} L(W; x, y) = \delta^3 (a^{(2)})^T$$

$$\Delta_{w^{1,3}} L(W; x, y) = \delta^3 (a^{(1)})^T$$

and for updating,

$$w^{1,3} = w^{1,3} - \eta \Delta_{w^{1,3}} L(w; x, y)$$

$$w^{2,3} = w^{2,3} - \eta \Delta_{w^{2,3}} L(w; x, y)$$

where η is learning rate.

$\delta^{(l)}$ is error term in l^{th} layer.

2. With each layer support increases from $n \times n$ to $(n+2) \times (n+2)$. In first layer it is 3×3. In second layer it will be 5×5. In third layer it will be 7×7. In fourth layer, it will be 9×9.

∴ Support of a neuron in 4^{th} non-image layer = $9 \times 9 = 81$ pixels

Scanned with CamScanner

3. We know that more complex models have lower bias and higher variance.

~~Ad~~ Increasing the number of hidden units will lead to more complex model. Hence it will lead to lower bias and higher variance.

Scanned with CamScanner

Q 4. $\tanh(x) = \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}} = \dfrac{1-e^{-2x}}{1+e^{-2x}}$ (Divide by e^{x} in numerator and denominator)

$$= \frac{1}{1+e^{-2x}} - \frac{(1+e^{-2x}-1)}{1+e^{-2x}} = \sigma(2x) - 1 + \frac{1}{1+e^{-2x}}$$

$$= 2\sigma(2x) - 1$$

$\therefore$ She is correct as there is a relation between $\sigma(x)$ and $\tanh(x)$.

$\therefore$ We have $\tanh\left(\frac{x}{2}\right) = 2\sigma(x) - 1$

$$\Rightarrow \sigma(x) = \frac{1+\tanh(x/2)}{2}$$

So we will substitute value of σ in y:

$$y_k = \sum_{j=1}^{M} w_{kj}^{(2)} \left(\frac{1}{2} + \frac{1}{2}\tanh\left(\frac{\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}}{2}\right)\right) + w_{k0}^{(2)}$$

$$= \sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2} \tanh\left(\frac{\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}}{2}\right) + w_{k0}^{(2)} + \sum_{j=1}^{M} \frac{w_{kj}^{(2)}}{2}$$

Hence we can write as –

$$y_k = \sum_{j=1}^{M} w'^{(2)}_{kj} \tanh\left(\sum_{j=1}^{D} w'^{(1)}_{ji} x_i + w'^{(1)}_{j0}\right) + w'^{(2)}_{k0}$$

where $w'^{(2)}_{kj} = \dfrac{w_{kj}^{(2)}}{2}$ for $j = 1$ to M

$w'^{(1)}_{ji} = \dfrac{w_{ji}^{(1)}}{2}$ for $j = 1$ to M and $i = 0$ to D.

$w'^{(2)}_{k0} = w_{k0}^{(2)} + \sum_{j=1}^{M} \dfrac{w_{kj}^{(2)}}{2}$

Hence,

Parameters of two models differ only by linear transformation.

Scanned with CamScanner

5. It is given that $H v_i = \lambda_i v_i$ — (i) where $\{v_i\}$ forms orthonormal set.

i.e. $\langle v_i, v_j \rangle = 1$ if $i = j$ — (ii)
$= 0$ otherwise

$$E(w) \approx E(w^*) + \frac{1}{2}(w - w^*)^T H (w - w^*)$$

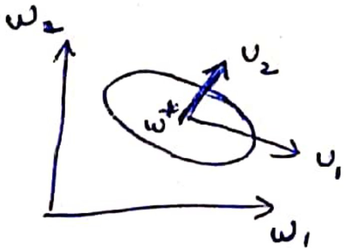
Putting $w - w^* = \sum_i \alpha_i v_i$

we have $E(w) \approx E(w^*) + \frac{1}{2}\left(\sum \alpha_i v_i\right)^T H \left(\sum_i \alpha_i v_i\right)$

$\therefore E(w) \approx E(w^*) + \frac{1}{2}\left(\sum_i \alpha_i v_i\right)^T \left(\sum_i \alpha_i \lambda_i v_i\right)$ {from (i)}

$\approx E(w^*) + \frac{1}{2} \sum_i \alpha_i^2 \lambda_i$ {from (ii)}

$\therefore$ Contours of constant error are ellipses whose axes are aligned with v_1, v_2 and length of axes are $\frac{1}{\sqrt{\lambda_1}}$ and $\frac{1}{\sqrt{\lambda_2}}$.

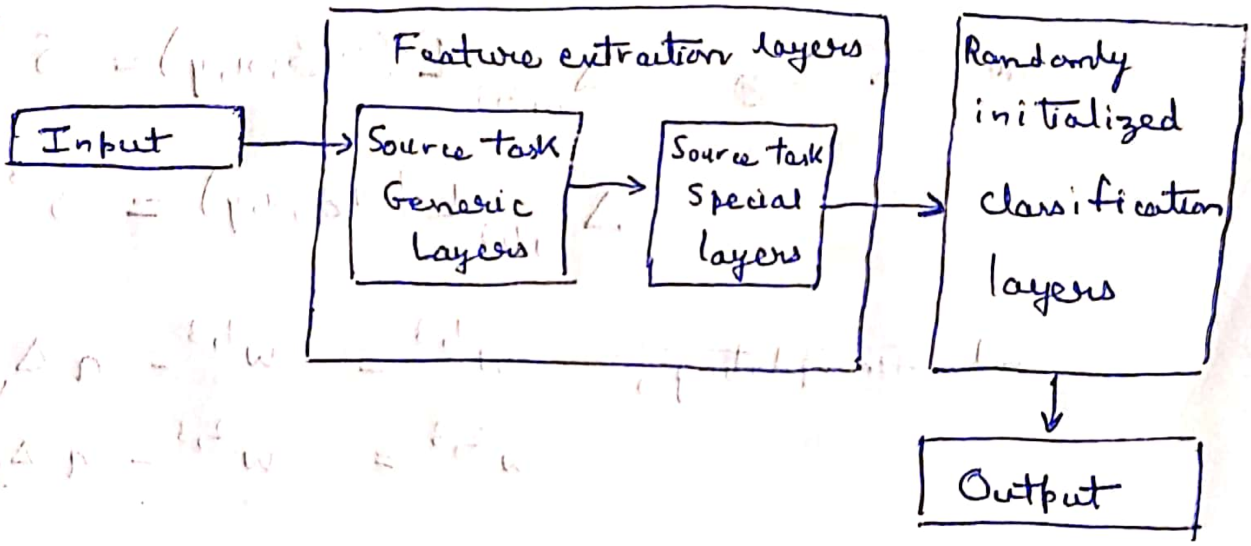


Ellipse is centered at w^*.

Scanned with CamScanner

6. We know that target and source datasets are similar because ~~one is of~~ & both includes images of animals from two different national parks.

~~So~~ Also for Kaziranga National Park, size of dataset is small (only 20 images of all 200 species as compared to 1 million images). So we will use following model:



In this model we use specialized features extractor of model deployed ~~to~~ at Olympic National Park as they are similar for source and target datasets. We randomly initialize parameters of classification layer and train them. Rest of the network remains frozen / unchanged in order to avoid overfitting.

Scanned with CamScanner