Assignment 3 (Sol 1,2): Deep Learning for Vision

Piyushi Manupriya

October 17, 2019

1. Answer to Question 1:

Notations

k	Depth of the Feed Forward Neural Network, not counting the input layer(layer 0) while computing depth.
d_k	No. of neurons in layer at depth k
$\mathbf{W}_k$	Weight matrix of dimension d_{k-1}xd_k. So $w_{ki,j}$ corresponds to weight of edge from node i in layer $k-1$ to node j in layer k. $\mathbf{w}_{kj}$ is a vector containing d_{k-1} weights from all neurons in layer $k-1$ to neuron j in layer k
n	No. of training points
o	No. of output neurons
$\tilde{y}_{j_i}$	Ground truth value for output neuron j given the sample i
y_{j_i}	Predicted value by output neuron j given the sample i
$g(\mathbf{x})$	$tanh(\mathbf{x})$, computed element-wise.

Expression for error:

$$E = \frac{1}{o.n}\sum_{j=1}^{o}\sum_{i=1}^{n}(\tilde{y}_{j_i} - y_{j_i})^2$$
$$y_{j_i} = g({\mathbf{w}_{kj}}^T g(\mathbf{W}_{k-1}^T g(...g(\mathbf{W}_3^T g(\mathbf{W}_2^T g(\mathbf{W}_1^T \mathbf{x}_i)))...))) \tag{1}$$

$$\frac{o.n}{2}\frac{\partial E}{\partial \mathbf{w}_{vj}} = -\sum_{i=1}^{n}(\tilde{y}_{ji} - y_{j_i})\frac{\partial y_{j_i}}{\partial \mathbf{w}_{vj}} \tag{2}$$

Since affine and tanh are differentiable functions and composition of differentiable functions is a differentiable function so y_{j_i} is a differentiable function of weights.

Let $\mathbf{l}_1 = \mathbf{W}_1^T \mathbf{x}_i$

$\mathbf{l}_i = \mathbf{W}_i^T g(\mathbf{l}_{i-1})$ or $\mathbf{l}_i = \sum_{j=1}^{d_i} {\mathbf{w}_{ij}}^T g(\mathbf{l}_{i-1})$ for i in range 2 to $k-1$

$l_k = \mathbf{w}_{kj}{}^T g(\mathbf{l}_{k-1})$.

So, $y_{j_i} = g(l_k)$.
Now using the above notations and applying chain rule with v lying in the range 1 to $k-1$:

$$\begin{aligned} \frac{\partial y_{j_i}}{\partial \mathbf{w}_{vj}} &= \frac{\partial \mathbf{l}_{vj}}{\partial \mathbf{w}_{vj}} \frac{\partial \mathbf{l}_{(v+1)j}}{\partial \mathbf{l}_{vj}} ... \frac{\partial l_{kj}}{\partial \mathbf{l}_{(k-1)j}} \frac{\partial y_{j_i}}{\partial l_{kj}} \\ \frac{\partial \mathbf{l}_{(v+1)j}}{\partial \mathbf{l}_{vj}} &= \frac{\partial g(\mathbf{l}_{vj})}{\partial \mathbf{l}_{vj}} \frac{\partial \mathbf{l}_{(v+1)_j}}{\partial g(\mathbf{l}_{vj})} \end{aligned} \tag{3}$$

Also,

$$\frac{\partial y_{j_i}}{\partial \mathbf{w}_{kj}} = \frac{\partial l_k}{\partial \mathbf{w}_{kj}} \frac{\partial y_{j_i}}{\partial l_k} = g(\mathbf{l}_{k-1}) \frac{\partial y_{j_i}}{\partial l_k} \tag{4}$$

At the origin in weight space, $\mathbf{l}_1$ is $\mathbf{0}$ and using the fact that $\tanh(0)$ is 0, we get $\mathbf{l}_{k-1} = \mathbf{0}$. Therefore using equation (4) we get,

$$\frac{\partial y_{j_i}}{\partial \mathbf{w}_{kj}} = 0 \tag{5}$$

Also, for v in range 1 to $k-1$, we have,

$$\frac{\partial \mathbf{l}_{(v+1)_j}}{\partial g(\mathbf{l}_{vj})} = \sum_{j=1}^{d_{v+1}} \mathbf{w}_{(v+1)_j} = 0 \tag{6}$$

Therefore, using equation (3) we get,

$$\frac{\partial y_{j_i}}{\partial \mathbf{w}_{vj}} = 0 \tag{7}$$

Now, for all v in range 1 to k, at origin in the weight space, we have $\frac{\partial y_{j_i}}{\partial \mathbf{w}_{vj}} = 0$, so from equation (2) $\frac{\partial E}{\partial \mathbf{w}_{vj}} = 0$ for all v in range 1 to k. Thus, proving that origin in the weight space is the stationary point of the error function.

2. Answer to Question 2:
Hessian Matrix:

$$\nabla^2_{\mathbf{w}} E = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \tag{8}$$

Let λ represent the eigenvalue of the matrix, characteristic equation for this matrix is

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda) = 0 \tag{9}$$

Hence, eigenvalues of the Hessian matrix are λ_1, λ_2