Q6) I Backpropagation Algorithm:

Initialize all weights to small random numbers. Until satisfied do

- For each training example, Do

i) Input the training example to the network and compute the network outputs

ii) For each output unit k

$$\delta_k \leftarrow o_k(1-o_k)(t_k-o_k)$$

iii) For each hidden unit h

$$\delta_h \leftarrow o_h(1-o_h)\sum w_{h,k}\delta_k$$

iv) Update each network weight $w_{i,j}$

$$w_{i,j} \leftarrow w_{i,j} + \Delta w_{i,j}$$

where, $\Delta w_{i,j} = \eta \delta_j x_{i,j}$

- Gradient descent over entire network weight vector
- Easily generalized to arbitrary directed graph
- Will find a local, not necessarily global error minimum.
- Often include weight momentum α

$$\Delta w_{i,j}(n) = \eta \delta_j x_{i,j} + \alpha \Delta w_{i,j}(n-1)$$

- Minimizes error over training examples
- Training can take thousands of iterations
- Using network after training is very fast.

II. Advanteges :- i) It is able to incorporate both direct evidence and indirect evidence into a single analysis.

ii) Able to produces results for all comparisons of interest within a connected network

iii) It is able to directly calculate the probability that each drug is the best treatment

iv) It is able to adjust for correlations within multi-arm traials

v) It is able to incorporate meta-regression to assess heterogeneity all within one model.

vi) Appears to produce valid, accurate results for the star and ladder network patterns

III. $\sin(x) + \sqrt{x^2 + y}$