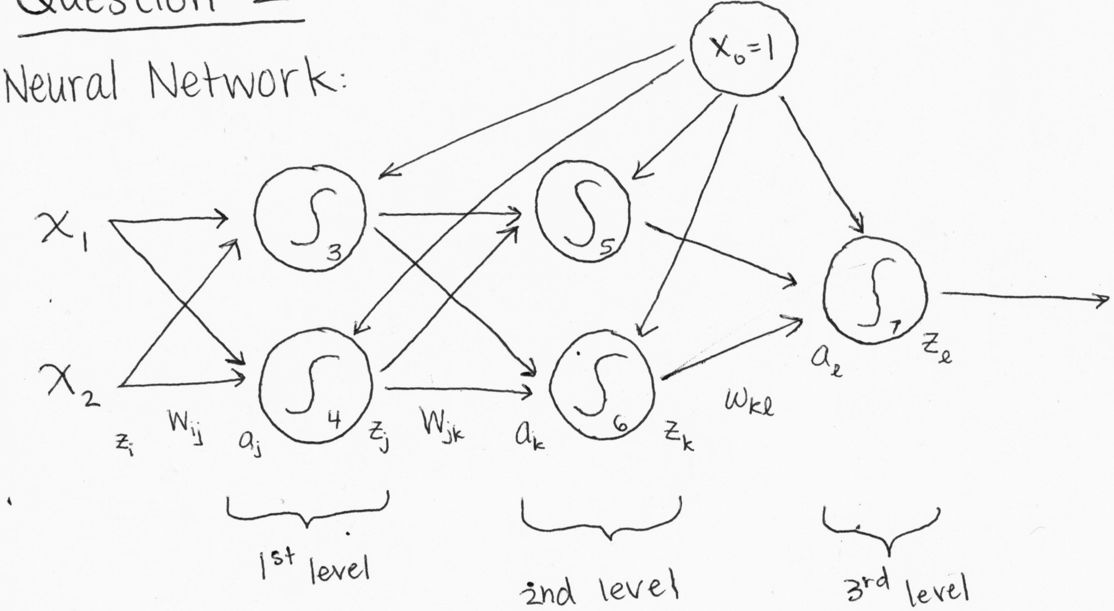
Question 2

Neural Network:



[gradient descent w/ logistic loss]

Logistic Loss function:

$$L(y, f(\vec{x};\theta)) = (y-1)\log(1-f(\vec{x};\theta)) - y\log(f(\vec{x};\theta))$$

where $f(\vec{x};\theta) = \dfrac{1}{1+\exp(-\theta\vec{x})}$

$$\frac{\partial R}{\partial w_{kl}} = \frac{1}{N}\sum_n \left[\frac{\partial L^n}{\partial a_l^n}\right]\left(\frac{\partial a_l^n}{\partial w_{kl}}\right) = \frac{1}{N}\sum_n \left[\frac{\partial (y-1)^n \log(1-f(a_l^n;\theta)) - y^n \log(f(a_l^n;\theta))}{\partial a_l^n}\right](Z_k^n)$$

where $f(x^n;\theta) = \dfrac{1}{1+\exp(-\theta x^n)}$

$$= \frac{1}{N}\sum_n \underbrace{\left[\frac{(y-1)^n \cdot f'(a_l^n;\theta)}{1-f(a_l^n;\theta)} - \frac{y^n \cdot f'(a_l^n;\theta)}{f(a_l^n;\theta)}\right]}_{\text{define as } \delta_l^n}(Z_k^n)$$

where $f'(a_l^n;\theta) = \dfrac{\theta\exp(\theta a_l^n)}{\exp(\theta a_l^n + 1)^2}$

& $f(a_l^n;\theta) = Z_l^n$

$$= \frac{\sum_N \delta_l^n Z_k^n}{N}$$ (output layer derivative)

Cost function: $= \dfrac{1}{N}\sum_n (y-1)^n \log\left(1 - f\left(\sum_k w_{kl} f\left(\sum_j W_{jk} f\left(\sum_i w_{ij} x_i^n\right)\right)\right)\right)$

$- y^n \log\left(f\left(\sum_k w_{kl} f\left(\sum_j w_{jk} f\left(\sum_i w_{ij} x_i^n\right)\right)\right)\right)$

$$\frac{\partial R}{\partial w_{jk}} = \frac{1}{N}\sum_n \left[\frac{\partial L^n}{\partial a_k^n}\right]\left[\frac{\partial a_k^n}{\partial w_{jk}}\right] = \frac{1}{N}\sum_n \left[\sum_l \frac{\partial L^n}{\partial a_l^n}\frac{\partial a_l^n}{\partial a_k^n}\right]\left(\frac{\partial a_k^n}{\partial w_{jk}}\right) = \frac{1}{N}\sum_n \left[\sum_l \delta_l^n \frac{\partial a_l^n}{\partial a_k^n}\right](z_j^n)$$

$$= \frac{1}{N}\sum_n \underbrace{\left[\sum_l \delta_l^n w_{kl} f'(a_k^n;\theta)\right]}_{\text{define as } \delta_k^n} z_j^n$$ (where f' as before)

$$= \frac{1}{N}\sum \delta_k^n z_j^n$$ (middle layer derivative)