2. (a) Our goal in linear regression is : $\min \sum_{i=1}^{n} e_i^2 \iff \min RSS$

$\iff \min \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} e_i = 0 \Rightarrow E(e_i) = 0$$

(b) from (a), we know that to minimize RSS,
we still need to : $\frac{\partial RSS}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$

$$\Rightarrow \sum_{i=1}^{n} x_i e_i = 0$$

(c) $\frac{1}{n}\sum(e_i - \bar{e})(\hat{Y}_i - \bar{\hat{Y}})$

$= \frac{1}{n}\sum(e_i \hat{Y}_i - \bar{e}\hat{Y}_i - e_i \bar{\hat{Y}} + \bar{e}\bar{\hat{Y}})$

$= \frac{1}{n}\sum[e_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - \bar{e}\hat{Y}_i] - \bar{e}\bar{\hat{Y}} + \bar{e}\bar{\hat{Y}}$

$= E(\hat{\beta}_0 e_i) + E(\hat{\beta}_1 x_i e_i)$

$= 0$

(d) from (a), we know that

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0$$

$$\Rightarrow E(Y_i) = E(\hat{Y}_i)$$

$$\Rightarrow \bar{\hat{Y}} = \bar{Y}$$

(e) $R^2 = 1 - \frac{RSS}{TSS}$

$$= 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2$$

$$= \sum_{i=1}^{n}(e_i + (\hat{Y}_i - \bar{Y}))^2$$

$\therefore TSS = RSS + ESS + 2\sum_{i=1}^{n} e_i(\hat{Y}_i - \bar{Y})$

$\sum e_i(\hat{Y}_i - \bar{Y}) = \sum e_i \hat{Y}_i - \sum e_i \bar{Y} = 0$

so $TSS = RSS + ESS$

$\therefore R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$

(f) $R^2 = \frac{ESS}{TSS}$

$S_{xx} = \sum (x_i - \bar{x})^2$

$S_{xY} = \sum (x_i - \bar{x})(Y_i - \bar{Y})$

$S_{YY} = \sum (Y_i - \bar{Y})^2$

we know that $\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}}$ $\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{x}$

so $\hat{Y}_i - \bar{Y} = \beta_1(x_i - \bar{x})$

$\therefore ESS = \sum \hat{\beta}_1^2 (x_i - \bar{x})^2 = \frac{S_{xY}^2}{S_{xx}}$

so $R^2 = \frac{S^2_{xY}}{S_{xx} S_{yy}} = \left(\frac{S_{xY}}{S_{xx} S_{YY}}\right)^2 = r^2_{xY}$

3. (a) $h_\theta(x_1) = \frac{1}{1+e^{-(0+0 \cdot 30)}} = \frac{1}{2}$

$h_\theta(x_2) = h_\theta(x_3) = h_\theta(x_4) = \frac{1}{2}$

(b) $\ell(\theta) = \sum_{i=1}^{4} [y^{(i)} \log(\frac{1}{2}) + (1-y^{(i)}) \log(\frac{1}{2})]$

$= 4\log(\frac{1}{2})$

(c) To maximize $\ell(\theta)$,

we need to write the deriviation:

$\frac{\partial \ell}{\partial \theta_0} = \sum_i (y_i - h_i) \quad \frac{\partial \ell}{\partial \theta_1} = \sum_i (y_i - h_i) x_i$

and then generate $\theta_j^{new} = \theta_j^{old} + \alpha \frac{\partial \ell}{\partial \theta_j}$ $(j = 0 \text{ or } 1)$

$\Rightarrow \frac{\partial \ell}{\partial \theta_0} = \sum_i (y_i - h_i) = 0$

$\frac{\partial \ell}{\partial \theta_1} = \sum_i (y_i - h_i) x_i = -0.5 \times 30 - 0.5 \times 50 + \frac{1}{2} \times 70 + \frac{1}{2} \times 90$

$= 40$

$\Rightarrow \theta_0^{(1)} = 0 + 0.01 \cdot 0 = 0 \quad \theta_1^{(1)} = 0 + 0.01 \times 40 = 0.4$

4. (a) $\hat{y}_1 = \arg\max_y w_y x$

$w_0 x_1 = 0.4 + 0.1 \times 3 - 0.3 \times 100 = -29.3$

$w_1 x_1 = 0.3 - 0.2 \times 3 + 0.5 \times 100 = 49.7 \quad \Rightarrow \hat{y}_1^{(0)} = 1 \neq 0$

$w_2 x_1 = -0.1 + 0.3 \times 3 + 0.2 \times 100 = 20.8$

so new $w^{(1)} = \begin{bmatrix} 1.4 & 3.1 & 99.7 \\ -0.7 & -3.2 & -99.5 \\ -0.1 & 0.3 & 0.2 \end{bmatrix} \Rightarrow \hat{y}_1^{(1)} = 0$

for x_2:

$w_0 x_2 = 1.4 \times 1 + 8 \times 3.1 + 300 \times 99.7 = 29936.2$

$w_1 x_2 = -29876.3 \quad \Rightarrow \hat{y}_2^{(1)} = 0 \neq 1$

$w_2 x_2 = 62.3$

$$\Rightarrow W^{(2)} = \begin{bmatrix} 0.4 & -4.9 & -200.3 \\ 0.3 & 4.8 & 200.5 \\ -0.1 & 0.3 & 0.2 \end{bmatrix} \Rightarrow \hat{y}_2^{(2)} = 1$$

for x_3 :

$w_0 x_3 = -30069.1$

$w_1 x_3 = 30099.3 \Rightarrow \hat{y}_3^{(2)} = 1 \neq 2$

$w_2 x_3 = 31.4$

$$\Rightarrow W^{(3)} = \begin{bmatrix} 0.4 & -4.9 & -200.3 \\ 0.7 & -0.2 & 50.5 \\ 0.9 & 5.3 & 150.2 \end{bmatrix}$$

(b) $x_4 = [1, 6, 200]$

$$W^{(3)} x_4 = \begin{bmatrix} -40089 \\ 10098.1 \\ 30072.7 \end{bmatrix} \qquad \hat{y}_4^{(3)} = \arg\max w_i^{(3)} x_4 = 2$$

$(i = 0, 1, 2)$