#1.

mymeasure(1000,50,100)
matmul: 0.00087308883667
my_mult: 4.32397603989
1.69858318973e-25

mymeasure(1000,1000,1000)
matmul: 0.0325238704681
my_mult: 898.295146942
4.13814283199e-20
[Finished in 902.8s]

#3.

(optimal value is highlighted in bold below)

M = 0; testing error = 66.1741164013; training error = 38.8665980131
M = 1; testing error = 56.7049816345; training error = 28.5347137491
M = 2; testing error = 35.7481373476; training error = 21.8938134083
M = 3; testing error = 39.9602812728; training error = 19.2864615083
M = 4; testing error = 5.38486882018; training error = 2.73776960577
M = 5; testing error = 5.39869662787; training error = 2.71673170645
M = 6; testing error = 6.6885168059; training error = 2.21265824391
M = 7; testing error = 6.92866621969; training error = 2.19889002467
M = 8; testing error = 6.86327749913; training error = 2.19732954624
M = 9; testing error = 79.3720539565; training error = 1.9429048648
M = 10; testing error = 133.671860772; training error = 1.91986460916
M = 11; testing error = 201.759217571; training error = 1.87475810502
M = 12; testing error = 173.934874322; training error = 1.87333639853
M = 13; testing error = 1733.969491; training error = 1.7945670175
M = 14; testing error = 3152977225.72; training error = 1.46188799934e-05
M = 15; testing error = 6865371022.83; training error = 2.61635899462e-06

#4.

(optimal value is highlighted in bold below, but while trying different values of alpha, I found at alpha = -5 fits the polynomial best. I believe there is an error in my training error, test error and validation error calculations)

alpha = -13; training error = 149.998440064; validation error = 122.775912571; test error = 173.909325821
alpha = -12; training error = 145.322095743; validation error = 123.710131343; test error = 169.821911096
alpha = -11; training error = 135.958442015; validation error = 124.187070257; test error = 161.560571606
alpha = -10; training error = 132.162800203; validation error = 112.534625599; test error = 152.156813384
alpha = -9; training error = 129.682933965; validation error = 95.9926304141; test error = 137.859090514
alpha = -8; training error = 130.502198171; validation error = 96.5870565638; test error = 131.502575838

alpha = -7; training error = 127.649483682; validation error = 100.594101755; test error = 126.334912998
alpha = -6; training error = 113.124004189; validation error = 98.4741648741; test error = 113.412013925
alpha = -5; training error = 86.3259796771; validation error = 74.8439697581; test error = 94.6179198541
alpha = -4; training error = 51.2970656163; validation error = 38.7152725039; test error = 64.3097959061
alpha = -3; training error = 22.0523781706; validation error = 17.6366339238; test error = 29.393171454
alpha = -2; training error = 16.3357087559; validation error = 23.7951844895; test error = 22.9004434438
alpha = -1; training error = 21.3614796762; validation error = 33.2623838459; test error = 26.685293809
alpha = 0; training error = 30.2578285171; validation error = 45.7307523313; test error = 33.5139536921
alpha = 1; training error = 29.932352322; validation error = 54.3476301526; test error = 34.2792136907
alpha = 2; training error = 34.972579795; validation error = 71.7954710998; test error = 40.8161441327

#5. I DON'T KNOW.

CSC411 Assignment 1

Akhil Gupta
1000357071

#2. Linear Regression : theory

$y(x) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M = w^T x$

$\ell(w) = \sum_n [t^{(n)} - y(x^{(n)})]^2$

$Z_{nm} = [x^{(n)}]^m$ $\qquad$ $\hat{y} = [y(x^{(1)}), y(x^{(2)}), \dots, y(x^{(N)})]$

(a) $\hat{y} = Zw$

$Z = N \times (M+1)$ matrix

$w = (M+1) \times 1$ vector ; $Zw = N \times (M+1) \times (M+1) \times 1$

$= \underline{N \times 1}$ vector

we know that $\hat{y}$ is a $N \times 1$ vector.

$$\hat{y} = \begin{bmatrix} y(x^{(1)}) \\ y(x^{(2)}) \\ \vdots \\ y(x^{(N)}) \end{bmatrix} = Zw = \begin{bmatrix} \sum_{i=0}^{M} w_i [x^{(1)}]^i \\ \sum_{i=0}^{M} w_i [x^{(2)}]^i \\ \vdots \\ \sum_{i=0}^{M} w_i [x^{(m)}]^i \end{bmatrix}$$

$\therefore y(x^{(N)}) = w_0 + w_1 [x^{(n)}]^1 + \dots + w_M [x^{(n)}]^M$

(b) $\ell(w) = \|Zw - t\|^2$

$= \sum_n (Zw - t)^2$ (by definition, $\|v\|^2 = \sum_n v_n^2$)

$= \sum_n (\hat{y} - t)^2$ (from (a))

$= \sum_n [y(x^{(n)}) - t^{(n)}]^2 = \ell(w)$

(c) $\frac{\partial l(w)}{\partial w_m} = -2\sum_n \left[t^{(n)} - y(x^{(n)}\right] \cdot Z_{nm}$

$l(w) = \sum_n \left[t^{(n)} - y(x^{(n)})\right]^2$

$\frac{\partial l(w)}{\partial w_m} = \frac{\partial}{\partial w_m} \sum_n \left[t^{(n)} - y(x^{(n)}\right]^2$

$= \sum_n \cdot 2\left[t^{(n)} - y(x^{(n)}\right] \cdot \frac{\partial}{\partial w_m}\left[t^{(n)} - y(x^{(n)})\right]$

$= 2\sum_n \left[t^{(n)} - y(x^{(n)}\right] \cdot \frac{\partial}{\partial w_m}\left[0 - \left(w_0 \cdot [x^{(n)}]^0 + w_1 \cdot [x^{(n)}]^1 + w_2 \cdot [x^{(n)}]^2 + \ldots + w_M \cdot [x^{(n)}]^M\right]\right.$

$= 2\sum_n \left[t^{(n)} - y(x^{(n)}\right] \cdot \left[-\underbrace{\left((x^{(n)})^0 + (x^{(n)})^1 + (x^{(n)})^2 + \ldots + (x^{(n)})^M\right.}_{Z_{nm}}\right]$

$= -2\sum_n \left[t^{(n)} - y(x^{(n)}\right] \cdot Z_{nm}$

(d) $\frac{\partial l(w)}{\partial w} = -2Z^T(t - \hat{y})$ $\qquad l(w) = \sum_n \left[t^{(n)} - y(x^{(n)})\right]^2$

for a vector v, $v^T v = \sum_i v_i^2$

$l(w) = \|Zw - t\|^2$

$l(w) = \|\hat{y} - t\|^2$

$\frac{\partial l(w)}{\partial w} = \sum_n (\hat{y} - t)^2$

$\frac{\partial l(w)}{\partial w} = 2\sum_n (\hat{y} - t) \cdot \frac{\partial(\hat{y} - t)}{\partial w}$

$= -2\sum_n (t - \hat{y}) \cdot \frac{\partial(\hat{y} - t)}{\partial w}$

$= -2\sum_n (t - \hat{y}) \cdot \frac{\partial(Zw - t)}{\partial w}$

$= -2\sum_n (t - \hat{y}) \cdot Z = -2Z\sum_n (t - \hat{y}) = -2Z^T(t - \hat{y})$

(e) $\frac{\partial l(w)}{\partial w} = 0,$

$Z^T Z$ is $(M+1) \times (M+1)$ matrix

$w = (Z^T Z)^{-1} Z^T t$

$y(x^{(n)}) = w^T x^{(n)}$

$= \begin{bmatrix} x^{(0)} \\ x^{(1)} \\ x^{(2)} \end{bmatrix} \cdot [w_0 w_1 \; \cdots \; w_m]$

$= w_0 x_0 + w_1 x_1 + w_2 x^2 + \ldots + w_m x^M$

$l(w) = \sum_n [t^{(n)} - y(x^{(n)})]^2$

$l(w) = (Zw - t)^T (Zw - t)$

$l(w) = ((Zw)^T - t^T)(Zw - t)$

$l(w) = (Zw)^T Zw - (Zw)^T t - t^T (Zw) + t^T t$

$l(w) = w^T Z^T Zw - 2(Zw)^T t + t^T t$

$\frac{\partial l(w)}{\partial w} = 2Z^T Zw - 2Z^T t = 0$

$Z^T Zw = Z^T y$

assuming $Z^T Z$ is invertible,

$w = (Z^T Z)^{-1} Z^T t$