Part.1
1). The learned weight vector is

1	-0.101137046
2	0.0458935299
3	-0.00273038670
4	3.07201340
5	-17.2254072
6	3.71125235
7	0.00715862492
8	-1.59900210
9	0.3736823375
10	-0.0157564197
11	-1.02417703
12	0.00969321451
13	-0.585969273
14	39.5843212

Note that the last one is from dummy variable

2) training ASE is 22.08127319
testing ASE is 22.6382563

3) Without the dummy variable, the training ASE is 24.47588278 and the testing ASE is 24.29223818. Both increased. The reason of this increment is that the 13 feathers are not the only feature to describe the housing price, so we need another variable, act as a constant value to describe the other features that will influence the housing price.

4) I added 1 to 50 random variable, from the graph I got from my program, the training ASE decrease while the testing ASE increase when adding more features, so more features didn't lead to better performance on prediction. The reason is when we adding more features, because we have more variable to describe the line, it gives us more precise regression of the training data, but there are many features actually not related to the housing price, so adding those features would instead hurting our performance on prediction.

Part2

1)
Training accuracy and testing accuracy increase and tend to be a fix number with the increase of iteration

2)
gradient

$$\underline{d} = \sum_{i=0}^{n} \left(y^i - g(\boldsymbol{w}^T)\right) x^i + \lambda w$$

Given training example (x^i, y^i), I = 1, …, N

Let w (0,0,0….,0)

Repeat until convergence encountered:
for i = 1 to N do:
        y = sigmoid(W*Xi)
        error = yi - y
        d = d + error * xi

w = w + n(d + lambda w)

3)
In my case, using modified algorithm didn't change the behavior a lot, but it do increase the maximum accuracy.