Answers to problem set #1

Supervised Learning: Regression and SVM

Data Mining, Spring 2018

1 Linear Regression

（1）在多元线性回归模型中：

Hypothesis: $h_\theta(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

Parameters: $\theta_0, \theta_1, \ldots, \theta_n$

Cost function:

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

利用梯度下降法求解参数的迭代过程如下：

Repeat {

$$\theta_j := \theta_j - \alpha \underbrace{\boxed{\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}}}_{\frac{\partial}{\partial \theta_0} J(\theta)}$$

(simultaneously update θ_j for $j = 0, \ldots, n$)

}

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \underbrace{x_0^{(i)}}_{x_0^{(i)} = 1}$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_1^{(i)}$$

$$\theta_2 := \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_2^{(i)}$$

$\ldots$

初始化：$\theta^0 = (0.00, 0.00, 0.00, 0.00, 0.00)^\mathsf{T}$

将学习率设为 1，迭代一次之后：

$$\theta^1 = (93.00, 8376.00, 6864.60, 8059.80, 8501.80)^\mathsf{T}.$$

（2）不能。因为

$$J(\theta^0) = 4328.50,\ J(\theta^1) = 3743077544562.61,\ J(\theta^0) < J(\theta^1).$$

（3）经一次迭代后计算$\Delta J = J(\theta^1) - J(\theta^0)$

α	$\boxed{\Delta J}$
0.1	-37407867726
0.01	-371787909.6
0.001	-3488802.333
0.0001	-11980.34696
1.00E-05	2170.964167
1.00E-06	250.7864054

因此取$\alpha = 0.00001$可使损失函数经一次迭代下降最快。

（4）利用标准方程求解参数，公式为:

$$\Theta = (X^T X)^{-1} X^T y$$

```
% linear regression (normal equation)
% Input: X, the feature matrix (m by (n+1))
%        y, the dependent variables
% Output: theta, the parameters
function theta = linear_regression_NE(X,y)
theta =  inv(X'*X)*X'*y;
end
```

求得：

$$\theta = (-19.50, 1.69, 0.38, -0.31, -0.44)^\mathsf{T}$$
$$m_{predict} = \theta^\mathsf{T} x = 89.51.$$

（5）引入正则化项之后，参数求解标准方程为：

$$\theta = [X^T X + \lambda \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}]^{-1} X^T Y$$

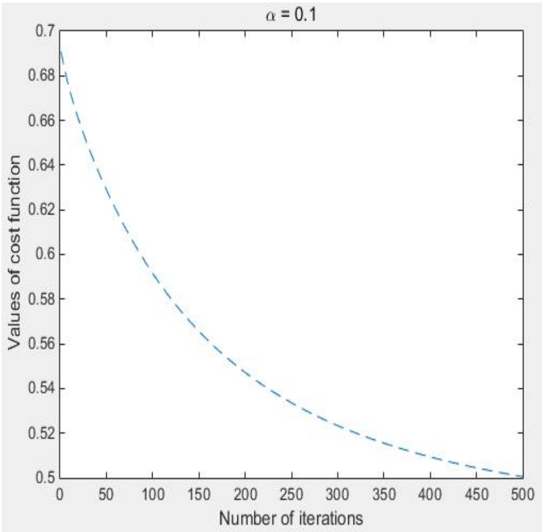
$$\lambda = 1, \theta = (-19.99, 1.47, 0.07, -0.23, -0.06)^{\mathsf{T}}$$

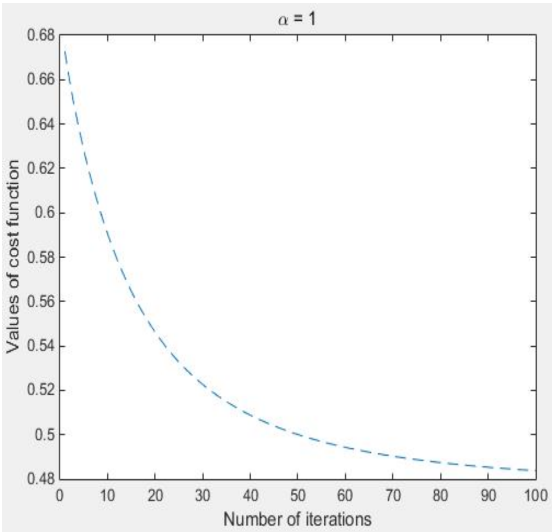
$$m_{predict} = \theta^{\mathsf{T}} x = 88.95.$$

（对预测结果简单分析，言之成理即可）。

2 Logistic Regression

（1）





$$\alpha = 1, \theta = (-2.67, 2.22, 1.06, -1.77, 2.24)^{\mathsf{T}}$$

%%%

```
% Call the library function glmfit
[theta,dev] = glmfit(X,y,'binomial','link','logit');
```

若调用 Matlab 库函数 glmfit 可得

$$\theta = (-2.67, 2.22, 1.06, -1.77, 2.24)^{\mathsf{T}}$$

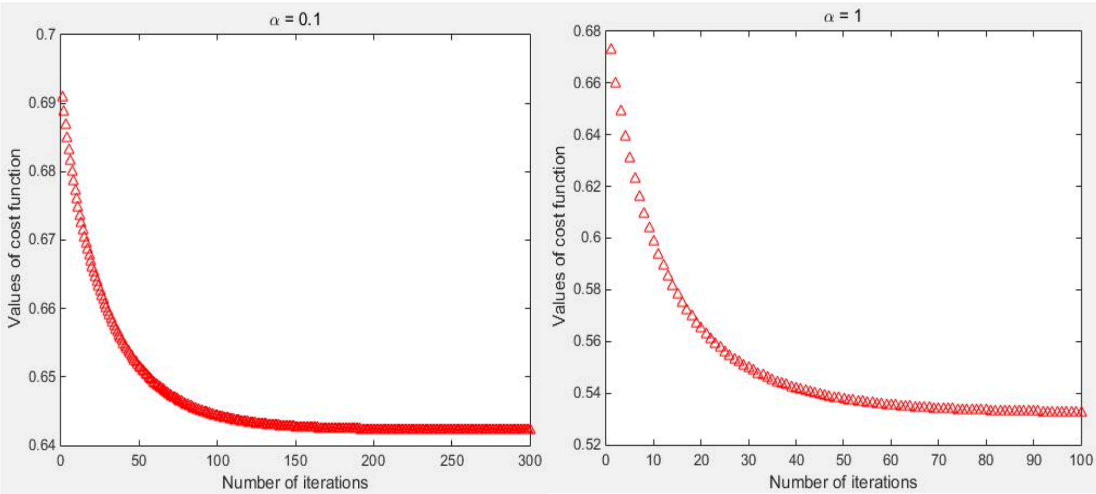
（2）

对影响子宫内膜癌发病的最直接的因素应为“雌激素药使用 ”及“非雌激素药使用 ”，因为通过逻辑回归模型求得的权重参数中这两项的值较大（其他答案言之成理即可）。

（3）

logistic_regression_GD_N.m

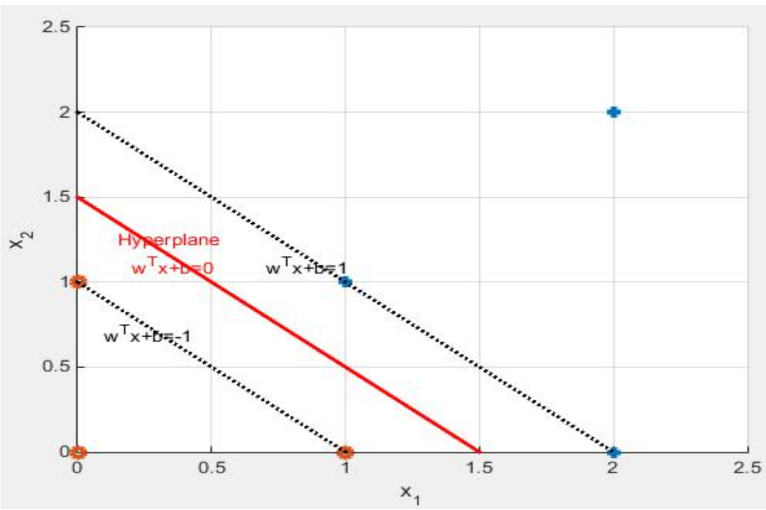
```
% logistic regression (gradient decent-- normalization)
% Input: X, the feature matrix
%        y, the depend variable
%        alpha, the learning rate
%        Iter, the number of iterations
%        lambda, the regulaization parameter
% Output: theta, the parameters
%         J, values of cost function
function [theta,J]= logistic_regression_GD_N (X,y,alpha,Iter,lambda)
```



$$\alpha = 1, \lambda = 1, \theta = (-1.00, 1.08, 0.51, -0.74, 0.81)^T$$

3 SVM

（1）支持向量与最优超平面如下图所示：



其中支持向量为：

$$(1,1)^{\mathsf{T}},(2,0)^{\mathsf{T}},(1,0)^{\mathsf{T}},(0,1)^{\mathsf{T}}$$

由于支持向量满足：

$$\begin{cases} \omega^{\mathsf{T}} x^{(i)} + b = +1, y^{(i)} = +1 \\ \omega^{\mathsf{T}} x^{(i)} + b = -1, y^{(i)} = -1 \end{cases}$$

将坐标点代入可得：

$$\omega = (2,2)^{\mathsf{T}}, b = -3$$

因此相应的超平面方程为：

$$2x_1 + 2x_2 - 3 = 0$$

（2）

由于新增的样本点能被正确分类且远离最优超平面，支持向量不发生变化，用 SVM 训练出来的分类模型不受影响；而逻辑回归模型的训练则要考虑所有训练样本点，因此新增训练样本点会影响逻辑回归（其他答案言之成理即可）。

（3）

由（1）可知支持向量为：

$$(1,1)^{\mathsf{T}},(2,0)^{\mathsf{T}},(1,0)^{\mathsf{T}},(0,1)^{\mathsf{T}}$$

两个异类支持向量到最优超平面（决策边界） 的距离之和即为“间隔”（margin）：

$$\gamma = \frac{2}{\|\omega\|} = \frac{\sqrt{2}}{2}$$

（4）

对拉格朗日函数求偏导后可得：

$$\omega = \Sigma_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

$$0 = \Sigma_{i=1}^{m} \alpha_i y^{(i)}$$

利用互补松弛性质有：

$$\alpha_i \geq 0$$

$$\alpha_i(y^{(i)}(\omega^{\mathsf{T}} x^{(i)} + b) - 1) = 0$$

$$\text{s.t.} \begin{cases} \alpha_i > 0, x^{(i)} \text{ is a support vector} \\ \alpha_i = 0, \text{otherwise} \end{cases}$$

代入支持向量坐标可得：

$$\omega = (\alpha_1 + 2\alpha_3 - \alpha_5, \alpha_1 - \alpha_6)^T$$

求解方程组：

$$\begin{cases} \alpha_1 + \alpha_3 - \alpha_5 - \alpha_6 = 0 \\ 2\alpha_1 + 2\alpha_3 - \alpha_5 - \alpha_6 + b = 1 \\ 2\alpha_1 + 4\alpha_3 - 2\alpha_5 + b = 1 \\ \alpha_1 + 2\alpha_3 - \alpha_5 + b = -1 \\ \alpha_1 - \alpha_6 + b = -1 \end{cases}$$

可得：

$$\begin{cases} \alpha_1 = 3 \\ \alpha_3 = 1 \\ \alpha_5 = 3 \\ \alpha_6 = 1 \\ b = -3 \end{cases}, \quad \alpha_2 = \alpha_4 = 0$$

超平面方程为：

$2x_1 + 2x_2 - 3 = 0$ 与（1）中的结论一致。