§ Machine Learning — Week 3 Assignment

1. Compute cost function & gradient

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]$$

$h_\theta(x) = g(\theta^T x) \leftarrow g(z) = \frac{1}{1+e^{-z}}$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

```
function [J, grad] = costFunction(theta, X, y)
%COSTFUNCTION Compute cost and gradient for logistic regression
%   J = COSTFUNCTION(theta, X, y) computes the cost of using theta as the
%   parameter for logistic regression and the gradient of the cost
%   w.r.t. to the parameters.

% Initialize some useful values
m = length(y); % number of training examples

% You need to return the following variables correctly
J = 0;
grad = zeros(size(theta));

% ====================== YOUR CODE HERE
======================
% Instructions: Compute the cost of a particular choice of theta.
%               You should set J to the cost.
%               Compute the partial derivatives and set grad to the partial
%               derivatives of the cost w.r.t. each parameter in theta
%
% Note: grad should have the same dimensions as theta


J= (1/m)*(-y'*log(sigmoid(X*theta))-(1-y)'*log(1-sigmoid(X*theta)));
```

1x100   100x1

```
grad=(1/m) *X'*(sigmoid(X*theta)-y);
```

3x100   100x1

```
%
```

==

end

2. predict

```
p=sigmoid(X*theta)>=0.5;
```

calculate the total number of data which is bigger than 0.5

→ belong to 1

Exercises 2. Regularized Logistic Regression

1. feature mapping

```
degree = 6;
out = ones(size(X1(:,1)));
for i = 1:degree
    for j = 0:i
        out(:, end+1) = (X1.^(i-j)).*(X2.^j);
end
end
```

$$\mathrm{mapFeature}(x) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1^3 \\ \vdots \\ x_1 x_2^5 \\ x_2^6 \end{bmatrix}$$

$$X = \begin{bmatrix} \begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix} & \underbrace{X_1 \quad X_2 \quad X_1^2 \quad \cdots \quad X_2^6}_{\text{features}} \end{bmatrix} \quad 100\times 28$$

2. cost function & gradient

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2.$$

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \qquad \text{for } j = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right) + \frac{\lambda}{m}\theta_j \qquad \text{for } j \ge 1$$

```
J = (sum(-y' * log(sigmoid(X * theta)) - (1 - y')*log(1 - sigmoid(X * theta))) / m)
+ lambda * sum(theta(2:end).^2) / (2*m);

grad =((sigmoid(X * theta) - y)' * X / m)' + lambda .* theta .* [0; ones(length(theta)-1,
1)] ./ m ;
```

pay attention that the grad is different when $y=0$. so we have to set $\theta_0=0$ to get the correct answer