Recap of Vectorized Form:

Cost Function

$$J(\theta) = -\frac{1}{m}\left[y^\top log(h) + (1-y)^\top h\right]$$

Gradient of Cost Function

$$\nabla J = \frac{\partial J(\theta)}{\partial \theta} = -\frac{1}{m}X^\top(y-h)$$

```
function [J, grad] = costFunction(theta, X, y)

    m = length(y); % number of training examples
    h = sigmoid(X*theta);

    % You need to return the following variables correctly
    J = -1/m * (y'* log(h) + (1-y)'*log(1-h));
    grad = -1/m * X'*(y-h);

end
```

Note:
X - rows are samples, columns are features.
y - column vector. Rows length are same as matrix X.
θ - column vector.

Regularized Cost Function

$$\begin{aligned} J_R(\theta) &= -\frac{1}{m}\sum_{i=1}^{m}\left[y^i log(h^i) + (1-y^i) log(1-h^i)\right] + \frac{\lambda}{m}\sum_{j=2}^{m}\theta_j^2 \\ &= J(\theta) + \frac{\lambda}{m}\sum_{j=2}^{m}\theta_j^2 \\ &= J(\theta) + \frac{\lambda}{2m}(\theta_{2:n}^\top \cdot \theta_{2:n}) \end{aligned}$$

Regularization index doesn't include bias term. So,

$$\begin{aligned} \frac{\partial J_R(\theta)}{\partial \theta_k} &= \frac{\partial J(\theta)}{\partial \theta_k} + \begin{cases} 0 & k = 1 \\ \frac{\lambda}{m}\theta_k & k \neq 1 \end{cases} \\ &= \frac{\partial J(\theta)}{\partial \theta} + \frac{\lambda}{m}\theta_{2:n} \end{aligned}$$

```
function [J, grad] = costFunctionReg(theta, X, y, lambda)

m = length(y); % number of training examples
h = sigmoid(X*theta);


J = -1/m * (y'* log(h) + (1-y)'*log(1-h)) +
        (lambda/(2*m))*theta(2:end)'*theta(2:end);

grad = -1/m * X'*(y-h) + [0;(lambda/m)*theta(2:end)];


end
```