

### Part 3 - Vector form of gradient for L2 penalized binary logistic regression

Derive the vector form of the first derivative of the L2-penalized  $J(\theta)$  with respect to  $\theta$ .

#### Solution

The gradient for this cost function is:

$$\nabla J(\theta) = -\frac{1}{m} \sum_{i=1}^m (y^{(i)} - h_{\theta}(x^{(i)})) x^{(i)} + \frac{\lambda}{m} \theta$$

---

Problem 1 continued on next page...

3

---

Statistical Machine Learning (Professor Devika Subramanian): Homework #2

Problem 1 (continued)

- $h_{\theta}(X)$  is an  $m$ -row prediction vector for all training examples, where  $X$  is matrix of training examples (each row is an example).
- $y$  is an  $m \times 1$  vector of true target values for training examples
- $X$  is matrix of all input features for training examples
- $\theta$  is parameter vector that has all model parameters

We can re-write as:

$$\nabla J(\theta) = \frac{1}{m} X^T (h_{\theta}(X) - y) + \frac{\lambda}{m} \theta$$