

1.4.

① $\nabla_{\theta}^2 \bar{J}_{\theta} = H = \frac{1}{m} (X^T S X + X \bar{I})$, show also H is positive def.

$$\nabla_{\theta} \bar{J}_{\theta} = \frac{1}{m} \left(\underbrace{\underbrace{1_m^T}_{\text{sum over column}} \underbrace{\text{diag}(h_{\theta} - y)}_{\substack{\text{row's } y \\ h_{\theta}(x_1) - y(x_1) \\ h_{\theta}(x_2) - y(x_2) \\ \vdots \\ h_{\theta}(x_m) - y(x_m)}}}_{m \times m} \right) \underbrace{X}_{m \times d}^T \quad \dots \quad (1 \times d \text{ shape})$$

Diagram illustrating the dimensions of the matrices in the gradient calculation:

- 1_m^T is a $1 \times m$ row vector.
- $\text{diag}(h_{\theta} - y)$ is an $m \times m$ diagonal matrix.
- X is an $m \times d$ matrix.
- The product $1_m^T \text{diag}(h_{\theta} - y)$ results in a $1 \times m$ row vector.
- Multiplying this row vector by X^T (which is $d \times m$) results in a $1 \times d$ row vector.

$$\begin{aligned} H(\theta) &= \nabla_{\theta} \left(\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_i \right) \\ &= \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) (1 - h_{\theta}(x^{(i)})) x^{(i)} x^{(i)T}) \\ &= \frac{1}{m} X^T \text{diag}(h_1(1-h_1), h_2(1-h_2), \dots, h_m(1-h_m)) X \\ &= \frac{1}{m} X^T S X \end{aligned}$$

②

To show H being positive def, assume for any nonzero \vec{v}

$$\vec{v}^T X^T S X \vec{v} = (\vec{v}^T X^T S^{\frac{1}{2}}) (S^{\frac{1}{2}} X \vec{v}) \quad (\because S \text{ diagonal})$$

$$= A^T A, \text{ where } A = S^{\frac{1}{2}} X \vec{v} \geq 0 \quad \#$$

1.5.

$$\Theta \leftarrow \Theta - H(\theta)^{-1} \nabla \bar{J}(\theta)$$