Non-invertibility (optional/advanced)

Suppose $m \leq n$

m: #examples, n: #features

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

$(X^T X)^{-1}$: non-invertible / singular

→ pinv: pseudo inverse

If $\lambda > 0$,

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \right)^{-1} X^T y$$

Regularized Logistic Regression

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

↓ new cost func^n with Regularization term

Grad Desc.

Repeat {

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

$$\theta_j := \theta_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} + \frac{\lambda}{m} \theta_j \right]$$

$\frac{\partial}{\partial \theta_j} J(\theta)$ ← new cost func^n

}