

# Advanced Optimisation for Logistic Regression.

Given  $\theta$ , we have code that can compute  
 $\rightarrow J(\theta)$

$$\rightarrow \frac{\partial}{\partial \theta_j} J(\theta) \text{ (for } j=0, 1, 2, \dots, n)$$

Optimization Algorithms:

- $\rightarrow$  Gradient descent (One we are using)
  - $\rightarrow$  Conjugate gradient
  - $\rightarrow$  BFGS
  - $\rightarrow$  L-BFGS
- Advantages:
- $\rightarrow$  No need to manually pick  $\alpha$
  - $\rightarrow$  Often faster than gradient descent

Disadvantage: More complex

---

Eg)  $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$  &  $J(\theta) = (\theta_1 - 5)^2 + (\theta_2 - 5)^2$

$$\text{for } \min_{\theta} J(\theta) \Rightarrow \theta_1 = 5, \theta_2 = 5$$

$$\frac{\partial}{\partial \theta_1} J(\theta) = 2(\theta_1 - 5) \quad \left\{ \quad \frac{\partial}{\partial \theta_2} J(\theta) = 2(\theta_2 - 5) \right.$$