

$y_i \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p(x_i; \theta)) = (p(x_i; \theta))^{y_i} (1 - p(x_i; \theta))^{1-y_i}$
 $p(y_1, \dots, y_n | x_1, \dots, x_n) = \prod_{i=1}^n (p(x_i; \theta))^{y_i} (1 - p(x_i; \theta))^{1-y_i}$
 $\log p(y_1, \dots, y_n | x_1, \dots, x_n) = \sum_{i=1}^n y_i \log p(x_i; \theta) + (1 - y_i) \log(1 - p(x_i; \theta))$
 $\max_{\theta} \log p(y_1, \dots, y_n | x_1, \dots, x_n) = \max_{\theta} \log p(y_1, \dots, y_n | x_1, \dots, x_n)$

$p_1 = p_0 + \Delta_1 \frac{\partial \log p(\cdot)}{\partial \beta}$
 $p_2 = p_1 + \Delta_2 \frac{\partial \log p(\cdot)}{\partial \beta}$

$\Delta_2 = \Delta_1 (0.9)^{\epsilon}$

$\ell(p) := \log p(y_1, \dots, y_n | x_1, \dots, x_n)$

$$\ell(p + \Delta) = \frac{\ell(p)}{0!} \Delta^0 + \frac{\ell'(p)}{1!} \Delta^1 + \frac{\ell''(p)}{2!} \Delta^2 + \dots$$

$$\ell(p + \Delta) \approx \ell(p) + \ell'(p) \Delta + \frac{\ell''(p)}{2} \Delta^2$$

$$\frac{\partial \ell(p + \Delta)}{\partial \Delta} \approx 0 + \ell'(p) + \frac{\partial \Delta}{\partial \Delta} \ell''(p) \Delta = 0$$

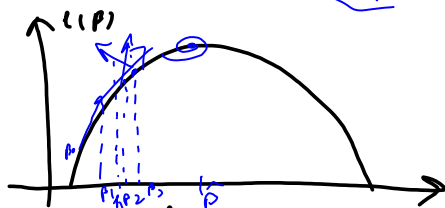
$$\Delta = \frac{-\ell'(p)}{\ell''(p)} \quad \Delta := \Delta(0.9)^{\epsilon}$$



much faster convergence than the first order optimization

Newton-Raphson optimization

$$p_i = p_{i-1} + \Delta_i \frac{\partial \ell(p_i)}{\partial \beta}$$



$S \subset \mathcal{D}$

$$\ell(p) = \log p(y_1, \dots, y_n | x_1, \dots, x_n) = \sum_{i=1}^n y_i \log p(x_i; p) + (1 - y_i) \log(1 - p(x_i; p))$$

$S \sim$ uniformly drawn $1 \dots n$
 $\|S\| \leq n$

$$\hat{\ell}(p) = \frac{n}{|S|} \sum_{i \in S} y_i \log p(x_i; p) + (1 - y_i) \log(1 - p(x_i; p))$$

$$\frac{\partial \hat{\ell}(p)}{\partial p} \text{ - unbiased for } \frac{\partial \ell(p)}{\partial p}$$