Derivation of $\nabla(-l_{(\theta)})$?

11.05.2018

$$l_{(\theta)} = -l_{(\theta)} = -\frac{1}{m}\sum_{j=1}^{m} y^{(j)} \log(h_\theta(x)) + (1-y^{(j)}) \log(1-h_\theta(x))$$

$$\frac{\partial l_{(\theta)}}{\partial \theta_i} = -\frac{1}{m}\sum_{j=1}^{m} y^{(j)} \frac{\frac{\partial}{\partial \theta_i}(h_\theta(x))}{h_\theta(x)} + (1-y^{(j)}) \frac{\frac{\partial}{\partial \theta_i}(1-h_\theta(x))}{(1-h_\theta(x))}$$

$$\underline{h_\theta(x) = g(\theta^T x)} = \frac{1}{1+e^{-\theta^T x}}$$

~~Let~~ task: $\frac{\partial}{\partial \theta_i} h_\theta(x)$; Let $u = 1+e^{-\theta^T x}$; $h_\theta(x) = \frac{1}{u}$

$$\frac{\partial u}{\partial \theta_i} = -e^{-\theta^T x} \cdot x_i$$

$$\partial h_\theta(x) = \frac{1}{u^2}$$

$$\frac{\partial}{\partial u} h_\theta(x) = \frac{-1}{u^2}$$

$$\frac{\partial}{\partial} h_\theta(x) = \frac{-1}{u^2} \times -e^{-\theta^T x} x_i \; \partial\theta_i$$

$$\frac{\partial}{\partial \theta_i} h_\theta(x) = \frac{e^{-\theta^T x}}{(1+e^{-\theta^T x})^2} x_i = \frac{1}{(1+e^{-\theta^T x})}\left(1 - \frac{1}{(1+e^{-\theta^T x})}\right) x_i$$

$$= g(\theta^T x)(1-g(\theta^T x)) x_i$$

$$\frac{\partial l(\theta)}{\partial \theta_i} = -\frac{1}{m}\sum_{j=1}^{m} y^{(j)} \frac{\cancel{g(\theta^T x)}(1-g(\theta^T x)) x_i}{\cancel{h_\theta(x)}} + (1-y^{(j)}) \frac{-(g(\theta^T x)\cancel{(1-g(\theta^T x))} x_i)}{\cancel{(1-h_\theta(x))}}$$

$$= \frac{1}{m}\sum_{j=1}^{m} \cancel{h_\theta(x)} (1-y^{(j)}) h_\theta(x^{(j)}) x_i^{(j)} - y^{(j)}(1-h_\theta(x^{(j)})) x_i^{(j)}$$

$$\boxed{\frac{\partial}{\partial \theta_i} l_{(\theta)} = \frac{1}{m}\sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)}) x_i^{(j)}}$$

$$\nabla(-l(\theta)) = \frac{1}{m}\begin{bmatrix} \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})\, x_0^{(j)} \\ \vdots \\ \sum_{j=1}^{m} (h_\theta(x^{(j)}) - y^{(j)})\, x_n^{(j)} \end{bmatrix}_{(n+1)\times 1}$$

Date:

* Logistic regression: *

11.05.2018

Random variable: $y|x;\theta \sim \text{Bernoulli}(\phi)$ — Prob. of Success

$P(y=1|x;\theta) = \phi$, $P(y=0|x;\theta) = 1-\phi$

hypothesis: $h_\theta(x) = \phi$

$$P(y|x;\theta) = \phi^y (1-\phi)^{1-y} = h_\theta(x)^y (1-h_\theta(x))^{1-y}$$

likelihood: $$L(\theta) = \prod_{j=1}^{m} P(y^{(j)}|x^{(j)};\theta)$$

$$= \prod_{j=1}^{m} h_\theta(x^{(j)})^{y^{(j)}} (1-h_\theta(x^{(j)}))^{1-y^{(j)}}$$

$$l(\theta) = \log L(\theta) = \frac{1}{m}\sum_{j=1}^{m} \log\left[h_\theta(x^{(j)})^{y^{(j)}} (1-h_\theta(x^{(j)}))^{1-y^{(j)}}\right]$$

$$= \frac{1}{m}\sum_{j=1}^{m} \left(y^{(j)} \log(h_\theta(x^{(j)})) + (1-y)^{(j)} \log(1-h_\theta(x^{(j)}))\right)$$

Convexity of log likelihood function?

$l(\theta)$ is Concave function.

Hence, negative log likelihood function is convex in nature. So $-l(\theta)$ has only one minimum (global minimum)

use gradient descent and newtons method to optimize / to find minimum of $-l(\theta)$.