On Optimizing a Logistic Regressor

Alexander G. Ororbia II

Introduction to Machine Learning

CSCI-335

3/20/2026

Special Thanks: Sargur N. Srihari, Andrew Ng, Carlos Guestrin

Gradient Descent

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

Goal: $\min_{\theta} J(\theta)$

Good news: Convex function!

Bad news: No analytical solution

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

What do we still need?

Derive gradient of $J(\theta_j)$ with respect to each θ_j

(Optimization!)

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

Derivation of gradient of J(Θ):

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

Gradient Descent

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

Goal: $\min_{\theta} J(\theta)$

***Good* news**: Convex function!

***Bad* news**: No analytical solution

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

}

(Simultaneously update all θ_j)

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)}$$

Gradient Descent

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

Goal: $\min_{\theta} J(\theta)$

Repeat { (***Simultaneously*** update all θ_j)

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

}

(Optimization!)

Gradient Descent for *Linear* Regression

Repeat {

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

$$h_\theta(x) = \theta^\top x$$

}

Gradient Descent for *Logistic* Regression

Repeat {

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

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

}

Gradient Descent for *Linear* Regression

Repeat {

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

}

$$h_\theta(x) = \theta^\top x$$

Woah, they're the same!!

Gradient D... *...ogistic* Regression

Repeat {

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

}

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

There Are More Forms of Regression at Your Fingertips

- Poisson regression $\lambda := \mathrm{E}(Y \mid x) = e^{\theta' x}$

$$p(y_1, \ldots, y_m \mid x_1, \ldots, x_m; \theta) = \prod_{i=1}^{m} \frac{e^{y_i \theta' x_i} e^{-e^{\theta' x_i}}}{y_i!}$$

- Exponential regression $\lambda_i = \exp(x_i{}'\beta)$

$$\prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i} \qquad l(\lambda) = n\ln(\lambda) - \lambda \sum x_i$$

- Other distributions and likelihoods/densities
  - You just need to be able to write out the density/likelihood, take the log-likelihood, and take derivatives of that log-likelihood with respect to your regression parameters Θ

Questions?

Deep robots!

Deep questions?!