Generalized Linear Models

Viswa Marepalli

May 2025

This document accompanies my article on GLMs, which can be found here.

Introduction

This document aims to have full derivations on some of the GLM lecture notes from CS229. The lecture notes I'm referencing from can be found here. This document will have minimal explanation. Refer to my article for full explanations (or even better, watch the CS229 lecture).

Bernoulli Distribution to Logistic Regression

Though I did write a fairly complete derivation on my article, it's here for completeness. In fairness, most of these derivations are just some logarithm rules.

$$p(y;\varphi) = \varphi^y(1-\varphi)^{1-y} \quad [1]$$

$$= \exp(\ln(\varphi^y(1-\varphi)^{1-y})) \quad [2]$$

$$= \exp(\ln(\varphi^y) + \ln((1-\varphi)^{1-y})) \quad [3]$$

$$= \exp(y\ln(\varphi) + (1-y)\ln(1-\varphi)) \quad [4]$$

$$= \exp(y\ln(\varphi) + \ln(1-\varphi) - y\ln(1-\varphi)) \quad [5]$$

$$= \exp(y(\ln(\varphi) - \ln(1-\varphi)) + \ln(1-\varphi)) \quad [6]$$

$$= \exp\left(y\ln\left(\frac{\varphi}{1-\varphi}\right) + \ln(1-\varphi)\right) \quad [7]$$

$$b(y) = 1 \quad [8]$$

$$\eta = \ln\left(\frac{\varphi}{1-\varphi}\right) \quad [9]$$

$$a(\eta) = -\ln(1-\varphi) \quad [10]$$

Now, we'll write $a(\eta)$ in terms of η rather than φ. We'll do this by re-arranging the definition we have for η for φ.

$$\eta = \ln\left(\frac{\varphi}{1-\varphi}\right) \quad [11]$$

$$e^\eta = e^{\ln\left(\frac{\varphi}{1-\varphi}\right)} = \frac{\varphi}{1-\varphi} \quad [12]$$

$$e^\eta(1-\varphi) = \varphi \quad [13]$$

$$e^\eta - \varphi e^\eta = \varphi \quad [14]$$

$$\varphi + \varphi e^\eta = e^\eta \quad [15]$$

$$\varphi(1+e^\eta) = e^\eta \quad [16]$$

$$\varphi = \frac{e^\eta}{1+e^\eta} = \frac{1}{1+e^{-\eta}} \quad [17]$$

Therefore:

$$a(\eta) = -\ln\left(1 - \frac{1}{1+e^{-\eta}}\right) \quad [18]$$

$$= -\ln\left(\frac{e^{-\eta}}{1+e^{-\eta}}\right) = -\ln\left(\frac{1}{1+e^\eta}\right) \quad [19]$$

$$= \ln(1+e^\eta) \quad [20]$$

Finding our hypothesis function now, using the fact that $E(y;\eta) = \frac{\partial}{\partial\eta}[a(\eta)]$:

$$\frac{\partial}{\partial\eta}(\ln(1+e^\eta)) = \frac{e^\eta}{1+e^\eta} = \frac{1}{1+e^{-\eta}} = \frac{1}{1+e^{-\theta^T x}} \quad [21]$$

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \quad [22]$$

Beautiful, right?

Gaussian Distribution to Linear Regression

This derivation assumes a variance $\sigma^2 = 1$.

$$p(y;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-\mu)^2\right) \quad [23]$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y^2 - 2\mu y + \mu^2)\right) \quad [24]$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2 + \mu y - \frac{1}{2}\mu^2\right) \quad [25]$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \exp\left(\mu y - \frac{1}{2}\mu^2\right) \quad [26]$$

$$b(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \quad [27]$$

$$\eta = \mu \quad [28]$$

$$a(\eta) = \frac{1}{2}\mu^2 \quad [29]$$

$$E[y;\mu] = \frac{\partial}{\partial \eta}\left(\frac{1}{2}\mu^2\right) \quad [30]$$

Substituting $\eta = \mu$:

$$E(y;\mu) = \frac{\partial}{\partial \eta}\left(\frac{1}{2}\eta^2\right) = \eta = \theta^T x \quad [31]$$

$$h_\theta(x) = \theta^T x \quad [32]$$

Those of you who've studied CS229, will know that this is the familiar hypothesis function that was introduced within the first set of lectures!