Motivation for the logistic function in logistic regression.

February 11, 2025

In this short note some additional motivation for the logistic function is provided. The idea is that we will posit a general model for the joint distribution of (y, x) and show that under this model the conditional probability $p(y = 1|\mathbf{x})$ has the form of a logistic function.

First we assume that the marginal distribution of y is $y \sim \text{Bernoulli}(\pi)$. Next, we assume that the conditional density of $\mathbf{x}|y = 1$ is of the form:

$$\begin{aligned} p(\mathbf{x}|y = 1) &= \exp(\boldsymbol{\theta}_1^\top \mathbf{x} - \kappa(\boldsymbol{\theta}_1))h(\mathbf{x}) \\ p(\mathbf{x}|y = 0) &= \exp(\boldsymbol{\theta}_0^\top \mathbf{x} - \kappa(\boldsymbol{\theta}_0))h(\mathbf{x}), \end{aligned}$$

where $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1$ are unknown parameters. Any random variable with this kind of density is said to come from an exponential family. A couple examples of this are:

1. $(p = 1)$ the Poisson distribution with parameter has a density/pmf of the form:

   $$p(x) = \frac{\lambda^x}{x!}\exp(-\lambda) = \exp(\log(\lambda)x - \lambda)\frac{1}{x!}.$$

   This is an exponential family of densities with parameter $\theta = \log(\lambda)$, $\kappa(\theta) = \lambda$ and $h(x) = (x!)^{-1}$.

2. The multivariate Gaussian distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ forms an exponential family. For simplicity let's show this in the case that $\boldsymbol{\Sigma} = \mathbf{I}_p$. The density can be written as

   $$\begin{aligned} p(\mathbf{x}) &= (2\pi)^{-p/2}\exp\big(-\frac{1}{2}\mathbf{x}^\top\mathbf{x} + \mathbf{x}^\top\mu - \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\mu}\big) \\ &= \exp\big(\boldsymbol{\mu}^\top\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\mu}\big)(2\pi)^{-p/2}\exp\big(-\frac{1}{2}\mathbf{x}^\top\mathbf{x}\big), \end{aligned}$$

   So $\boldsymbol{\theta} = \boldsymbol{\mu}$, $\kappa(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\mu}$, and $h(\mathbf{x}) = (2\pi)^{-p/2}\exp\big(-\frac{1}{2}\mathbf{x}^\top\mathbf{x}\big)$.

Exponential family distributions are ubiquitous and cover a wide variety of potential conditional distributions for the feature vector $\mathbf{x}$.

Let's return to the joint model of $(y, \mathbf{x})$ and compute $p(y = 1|\mathbf{x})$:

$$\begin{aligned} p(y = 1|\mathbf{x}) = \frac{p(y = 1, \mathbf{x})}{p(\mathbf{x})} &= \frac{p(\mathbf{x}|y = 1)p(y = 1)}{p(\mathbf{x}|y = 1)p(y = 1) + p(\mathbf{x}|y = 0)p(y = 0)} \\ &= \frac{\exp(\boldsymbol{\theta}_1^\top\mathbf{x} - \kappa(\boldsymbol{\theta}_1))h(\mathbf{x})\pi}{\exp(\boldsymbol{\theta}_1^\top\mathbf{x} - \kappa(\boldsymbol{\theta}_1))h(\mathbf{x})\pi + \exp(\boldsymbol{\theta}_0^\top\mathbf{x} - \kappa(\boldsymbol{\theta}_0))h(\mathbf{x})(1 - \pi)}. \end{aligned}$$

Canceling the $h(\mathbf{x})$ terms and dividing the denominator by the numerator we get:

$$\begin{aligned}
p(y=1|\mathbf{x}) &= \frac{1}{1+\exp\big((\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^\top\mathbf{x}+\kappa(\boldsymbol{\theta}_1)-\kappa(\boldsymbol{\theta}_0)+\log(\frac{1-\pi}{\pi})\big)} \\
&= \sigma\big((\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)^\top\mathbf{x}+\kappa(\boldsymbol{\theta}_0)-\kappa(\boldsymbol{\theta}_1)+\log(\frac{\pi}{1-\pi})\big).
\end{aligned}$$

Consequently, if we call $\boldsymbol{\beta}=\boldsymbol{\theta}_1-\boldsymbol{\theta}_1$ and $\beta_0=\kappa(\boldsymbol{\theta}_0)-\kappa(\boldsymbol{\theta}_1)+\log(\frac{\pi}{1-\pi})$, the above expression can be written as $\sigma(\boldsymbol{\beta}^\top\mathbf{x}+\beta_0)$. Thus, the logistic regression model encapsulates any model for the joint distribution of $(y,\mathbf{x})$ where the conditional distribution of $\mathbf{x}$ is from an exponential family. Note that if we were willing to make more assumptions about the form of the exponential family distribution from which $\mathbf{x}$ is derived this would potentially place additional constraints on $\boldsymbol{\beta}$ and β_0.