Generalized Linear Models

Closely related to generalized linear models (GLMs) are exponential families. An exponential family is one whose probability distribution function can be written as

$$P(y;\eta) = b(y)\exp\left(\eta^T T(y) - a(\eta)\right)$$

where y is the data, η is the natural parameter, $T(y)$ is a sufficient statistic, $b(y)$ is a base measure, and $a(\eta)$ is a log-partition function. As long as all functions are only dependent on the input variable, there can be any choice of definitions for T, b, and a as long as the distribution function integrates to 1.

An example of a member of the exponential family is the Bernoulli distribution, which has a probability distribution function of

$$P(y;\phi) = \phi^y(1-\phi)^{(1-y)}$$

which can then be rewritten as

$$\exp\left(\log\left(\phi^y(1-\phi)^{(1-y)}\right)\right)$$

which breaks into

$$\exp\left(y\log\phi + (1-y)\log\left(1-\phi\right)\right)$$

and then

$$\exp\left(\log\left(\frac{\phi}{1-\phi}\right)y + \log\left(1-\phi\right)\right)$$

which allows us to see that $b(y) = 1$, $\eta^T = \eta = \log\frac{\phi}{1-\phi}$, $T(y) = y$, and $a(\eta) = -\log\left(1-\phi\right)$. Since we want an expression in temrs of η for $a(\eta)$, we can solve for it using $\eta = \log\frac{\phi}{1-\phi}$ which results in

$$\phi = \frac{1}{1+e^{-\eta}}$$

so therefore

$$a(\eta) = -\log\left(1 - \frac{1}{1+e^{-\eta}}\right) = \log\left(1+e^{\eta}\right)$$

Some properties of exponential families are

1. the MLE (maximum likelihood estimation) w.r.t η is concave
2. the mean $\mathrm{E}[y;\eta]$ is equivalent to $\frac{\partial}{\partial\eta}a(\eta)$
3. the variance $\mathrm{Var}[y;\eta]$ is equivalent to $\frac{\partial^2}{\partial\eta^2}a(\eta)$

Assume that $y \mid x; \theta \sim \text{Exponential Family}(\eta)$ and $\eta = \theta^T x$. At test time, the output will be $\mathrm{E}\left[y \mid x; \theta\right]$, which means that our hypothesis function is $h_\theta(x) = \mathrm{E}\left[y \mid x; \theta\right] = \frac{\partial}{\partial \eta} a(\eta)$.

Going back to Bernoulli's distribution, we see that

$$\mathrm{E}\left[y \mid x; \theta\right] = \frac{e^\eta}{1 + e^\eta} = \frac{1}{1 + e^{-\eta}}$$

which is equivalent to

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

which is the sigmoid function used in logistic regression. We can say that logistic regression is the natural result of choosing the Bernoulli distribution for a generalized linear model.