ML Fundamentals: Generalized Linear Models

February 2, 2021

1 Generalized Linear Models

We know the probability distribution in the regression and classification settings.
Linear Regression:

$$y \mid x; \theta \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Classification (Logistic Regression):

$$y \mid x; \theta \sim \text{Bernoulli}(\phi)$$

Lets define GLM (Generalised Linear Models) family, where the distribution can be written in the form as:

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

Where η is a natural parameter, and $T(y)$ is sufficient statistics, and $a(\eta)$ is called log partition function.

1.1 Bernoulli Distribution as Generalised Linear Model

Bernoulli specify distribution over $y \in \{0, 1\}$

$$p(y = 1; \phi) = \phi$$

$$p(y = 0; \phi) = 1 - \phi$$

$$\begin{aligned} p(y; \phi) &= \phi^y (1 - \phi)^{1-y} \\ &= \exp(y \log \phi + (1 - y) \log(1 - \phi)) \\ &= \exp\left(\left(\log\left(\frac{\phi}{1 - \phi}\right)\right) y + \log(1 - \phi)\right) \end{aligned}$$

Thus the natural parameter can be given as $\eta = \log(\phi/(1 - \phi))$ Hence we can derive,

$$\begin{aligned} T(y) &= y \\ a(\eta) &= -\log(1 - \phi) \\ &= \log\left(1 + e^\eta\right) \\ b(y) &= 1 \end{aligned}$$

Hence, we have shown, Bernoulli distribution can be written as GLM.

1.2 Gaussian Distribution as Generalised Linear Model

Lets set $\sigma^2 = 1$

$$\begin{aligned} p(y; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y^2\right) \cdot \exp\left(\mu y - \frac{1}{2}\mu^2\right) \end{aligned}$$

Thus we see Gaussian is also in exponential family.

$$\begin{aligned}
\eta &= \mu \\
T(y) &= y \\
a(\eta) &= \mu^2/2 \\
&= \eta^2/2 \\
b(y) &= (1/\sqrt{2\pi}) \exp\left(-y^2/2\right)
\end{aligned}$$

Other distribution like following are also member of exponential family

- Multinomial distribution
- Poisson distribution to model count-data
- the Gamma and the Exponential (for modelling continuous, non-negative random variables, such as time intervals)
- the Beta and the Dirichlet (for distributions over probabilities)

1.3 Constructing GLMs

Consider a classification or regression problem where we would like to predict the value of some random variable y as a function of x.
We will make following three examples.

- $y \mid x; \theta \sim$ ExponentialFamily (η) Given x and θ, y follows exponential distribution with parameter η
- We need to come up with hypothesis, $h(x) = \mathrm{E}[y \mid x]$, that given x and θ, it should give expected value of y.
- Lets make a design choice, $\eta = \theta^T x$

1.3.1 Constructing GLMs: Ordinary Least Squares

Lets consider the case for linear regression. Where noise follows the Gaussian distribution, hence value of y given x and θ are from Gaussian distribution. Gaussian distribution, the expected value of y is μ, and $\eta = \mu$ as per the link between Gaussian distribution and GLM. Further, $\eta = \theta^T x$ was the design choice.
Hence,

$$\begin{aligned}
h_\theta(x) &= E[y \mid x; \theta] \\
&= \mu \\
&= \eta \\
&= \theta^T x
\end{aligned}$$

1.3.2 Constructing GLMs: Logistic Regression

Logistic regression, we know, y is binary valued, given x, it is expected to be class 1 with probability ϕ. As per Bernoulli distribution, the link between Bernoulli distribution's canonical parameter ϕ with GLM natural parameter is $\phi = 1/\left(1 + e^{-\eta}\right)$.

Hence,

$$
\begin{aligned}
h_\theta(x) &= E[y \mid x; \theta] \\
&= \phi \\
&= 1/\left(1 + e^{-\eta}\right) \\
&= 1/\left(1 + e^{-\theta^T x}\right)
\end{aligned}
$$