CS229 Machine Learning Note 3

Generalized linear models (GLM)

A broader family of models for previous cases.

3.1 The Exponential Family

Exponential family, defined as

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

η: **Natural parameter** (also **Canonical parameter**) of the distribution.
$T(y)$: **Sufficient Statistic** for the distribution.
$a(\eta)$: **log partition function**, $e^{-a(\eta)}$ plays the role of a normalization constant that ensure integration over y to 1.

Fixed choice of T a and b defines a family of distributions that is parameterized by η. (Only differed by η)

Example on bernoulli distribution

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

$$\implies \eta = \log(\frac{\phi}{1-\phi}) \implies \phi = \frac{1}{1+e^{-\eta}}$$

$$T(y) = y, \quad a(\eta) = \log(1+e^{\eta}), \quad b(y) = 1$$

Example of Normal Distribution (Here $\sigma = 1$ for simplification)

$$
\begin{aligned}
p(y;\mu) &= \frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}(y-\mu)^2\right) \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)\exp\left(\mu y-\frac{1}{2}\mu^2\right) \\
\implies \eta = \mu, \quad T(y) = y, \quad a(\eta) = \frac{\eta^2}{2}, \quad b(y) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)
\end{aligned}
$$

> *Note:* ***A more general definition of exponential family***
>
> $$p(y;\eta,\tau) = b(a,\tau)\exp(\frac{\eta^T T(y) - a(\eta)}{c(\tau)})$$
>
> where τ is called **dispersion parameter**, σ^2 for $\mathcal{N}$

Other distributions in exponential family:

- Poisson, Multinomial, Gamma, Exponential, Beta, Dirichlet.

3.2 Constructing GLM

Usually, make following assumptions:

1. $y|x;\theta \sim \text{ExponentialFamily}(\eta)$
2. Goal is to **predict expected** $T(y)$, mostly, $y \implies$ hypothesis h satisfy $h(x) = E[y|x]$
3. Natural parameter η and inputs x are linearly related: $\eta = \theta^T x$ (or $\eta_i = \theta_i^T x$ for vec)

3.2.1 Ordinary least squares

y: **Target variable** or **Response variable**, $Y|X \sim \mathcal{N}(\mu, \sigma^2)$

$$h_\theta(x) = E[y|x;\theta] = \mu = \eta = \theta^T x$$

3.2.2 Logistic regression

$$h_\theta(x) = E[y|x;\theta] = \phi = \frac{1}{1+e^{-\eta}} = \frac{1}{1+e^{-\theta^T x}}$$

(The forth equation is based on assumption 3)

3.2.3 More terminologies

Canonical response function: $g(\eta) = E[T(y); \eta]$
Canonical link function g^{-1}, the inverse of above.

For Gaussian family, g is identify function; For Bernoulli, logistic function