Generalised linear models (ENCODING & DECODING II)

- In order to have a full picture of GLMs, we need to talk first about the exponential family of probability distributions:

- Any prob. distribution of the form

$$p(y \mid \vec{\eta}(\vec{\theta})) = h(y) \exp\{\vec{\eta}(\vec{\theta}) \cdot \vec{t}(y) - A(\vec{\eta}(\vec{\theta}))\}$$

$$\begin{cases} \vec{\theta} : \text{parameters} \\ h(y) : \text{Base} \\ \vec{\eta}(\vec{\theta}) : \text{natural parameters} \\ \vec{t}(y) : \text{sufficient statistics} \\ A(\vec{\eta}) : \text{log partition} \end{cases}$$

- Poisson distribution

$$p(y \mid \lambda) = \frac{\lambda^y}{y!} e^{-\lambda} = \frac{1}{y!} e^{y \log \lambda - \lambda}$$

$$\rightarrow \begin{cases} \frac{1}{y!} = h(y) \end{cases} \qquad \log \lambda = \eta \qquad A(\eta) = e^{\eta}$$

$$A(\lambda) = \lambda \qquad \rightarrow p(y|\eta) = \frac{1}{y!} e^{(\eta y - e^{\eta})}$$

- Bernoulli

$$p(y \mid \pi) = \cancel{} \pi^y (1-\pi)^{1-y} \qquad (y \in \{0,1\})$$

$$= e^{y \ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)}$$

$$\begin{cases} h(y) = 1 \\ \eta = \log\left(\frac{\pi}{1-\pi}\right) \end{cases} \qquad A(\pi) = -\ln(1-\pi) \qquad \pi = \frac{e^{\eta}}{1+e^{\eta}} = \frac{1}{1+e^{-\eta}}$$

$$A(\eta) = \ln(1+e^{\eta})$$

$$\rightarrow p(y|\eta) = e^{\eta y - \ln(1+e^{\eta})}$$

. Exponential family distributions

- Poisson
- Bernoulli
- Gaussian
- Gamma
- exponential

- Non-exponential :
  - Uniform
  - t-student
  - ε

- Property of exponential families

- $\boxed{\vec{\nabla}_\eta A(\vec{\eta}) = \mathbb{E}[\vec{t}(y)]}$ (Exercise ?)

Poisson : $\begin{cases} A = e^{\eta} \rightsquigarrow A' = e^{\eta} = e^{\log \lambda} = \lambda \\ \mathbb{E}[y] = \lambda \end{cases}$

Bernoulli $\begin{cases} A = \ln(1+e^{\eta}) \rightsquigarrow A' = \dfrac{e^{\eta}}{1+e^{\eta}} = \pi \\ \mathbb{E}[y] = \pi \end{cases}$

GLMs

~~[illegible]~~ We have: $\{\vec{x}_1, \dots, \vec{x}_N\}$ and $\{y_1 \dots y_N\}$
N datapoints

- Encoding model where we assume in general that:

$$p(y_i \mid \vec{x}_i) = h(y_i) \exp\{\eta_i y_i - A(\eta_i)\}$$

where:

$$\begin{cases} \eta = \psi(\mu) & \leftarrow \left(\begin{array}{l} \text{Poisson: } \eta = \log\mu \\ \text{Binomial: } \eta = \log\left(\frac{\mu}{\mu-1}\right) \end{array} \right) \\ \mu_i = f(\vec{w}\vec{x}_i) & \leftarrow \boxed{f \text{ is the link function}} \\ \mu = \mathbb{E}[y] & \end{cases}$$

- Remember that $\nabla_\eta A(\eta) = \mathbb{E}[\vec{t}(y)] \Rightarrow \frac{dA}{d\eta} = \mathbb{E}[y] = \mu$

$$\Rightarrow \eta = \psi(\mu),\ \mu = \psi^{-1}(\eta)$$

$\boxed{\Rightarrow \text{There is always a 1-1 map between } \eta \text{ and } \mathbb{E}[y]}$

$$p(y_i \mid \vec{x}_i) = h(y_i) \exp\{\psi(f(\vec{w}\vec{x}))y_i - A(\psi(f(\vec{w}\vec{x}_i)))\}$$

If ~~[illegible]~~ $f = \psi^{-1} \Rightarrow f$ is the canonical link function

$$p(y_i \mid \vec{x}_i) = h(y_i) \exp\{(\vec{w}^T\vec{x}_i)y_i - A(\vec{w}\vec{x}_i)\}$$

(Poisson can. link $f = e^x$
Binomial can. link $f = \frac{1}{1+e^{-x}}$ sigmoid)

Model	Noise	Canonical Link
linear regression	Gaussian (known σ^2)	identity
Poisson regression	Poisson	log
logistic regression	Bernoulli	log-odds ($\ln(\frac{\mu}{1-\mu})$)

(Mention I don't have time here but linear regression is an interesting example to analyze)

- If firing rate is large → Gaussian noise
  is close to 0 → Poisson

  If $y \in \{0,1\}$ → logistic regression

Fitting a GLM

- In general $\vec{w}_{t+1} = \vec{w}_t + \alpha \nabla_{\vec{w}} \mathcal{L}(\vec{y} \mid X, \vec{w})$

gradient ~~descent~~ ascent ← Maximize the likelihood
(If minimize the loss, then we would use grad. descent)

Therefore: $\nabla_{\vec{w}} \mathcal{L}(\vec{y} \mid X, \vec{w})$ in GLMs?

$$\mathcal{L}(\vec{y} \mid X, \vec{w}) = \sum_{i=1}^{N} \mathcal{L}_i = \cancel{\sum_{i=1}^{N}\left(\log(h(y_i)) + \eta_i y_i - A(\eta_i)\right)}$$

$$\nabla_{\vec{w}} \mathcal{L} = \sum_{i=1}^{N} \left((\nabla_{\eta_i} \mathcal{L}_i)(\nabla_{\vec{w}} \eta_i) \right) = \sum_{i=1}^{N} \left((y_i - \nabla_\eta A)(\nabla_w \eta_i) \right) = *$$

$\nabla_\eta A = \mathbb{E}[y]$ $\quad \eta = \psi(\mu)$ $\quad \mu = f(\vec{w}\vec{x})$ $\quad \nabla_w \vec{w}\vec{x}_i = \vec{x}_i$

$$= \sum_{i=1}^{N} \left((y_i - \mathbb{E}[y_i]) \underbrace{(\vec{\nabla}_{\mu_i} \eta_i)}_{\psi'} \underbrace{(\vec{\nabla}_{\vec{w}\vec{x}_i} \mu_i)}_{f'} \vec{x}_i \right)$$

$*$ If canonical form $\eta = \psi(f(\vec{w}\vec{x})) = \vec{w}\vec{x}$

$$= \sum_{i=1}^{N} \left((y_i - f(\vec{w}\vec{x})) \vec{x}_i \right)$$

$$\left(\begin{array}{l} \vec{\nabla}_w \eta_i = \\ \vec{\nabla}_w (\vec{w}\vec{x}_i) = \vec{x}_i \end{array} \right)$$