$$p(\bar{x}|\bar{\theta}) = h(\bar{x})\, g(\bar{\theta})\, e^{\eta(\bar{\theta})^T \bar{t}(\bar{x})} =$$

$$= h(\bar{x})\, e^{\bar{\eta}(\bar{\theta})^T \bar{t}(\bar{x}) - a(\bar{\theta})}$$

$$\bar{\mu} = \mathbb{E}_{p(\bar{x}|\bar{\theta})}[\bar{t}(\bar{x})] = \nabla_{\bar{\theta}}\, a(\bar{\theta})$$

$$\mathrm{Var}_p[\bar{t}(\bar{x})] = H(a(\bar{\theta})) = \left(\frac{\partial^2 a}{\partial\theta_i \partial\theta_j}\right)_{ij}$$

---

linear regr.: $\hat{y} \sim \bar{x}^T \bar{w} + \varepsilon$

$\hat{y} = \bar{w}^T \bar{x}$

classification

logistic regr.: $\hat{y} = \sigma(\bar{w}^T \bar{x})$

$\hat{y} = h(\bar{w}^T \bar{x})$

Generalized linear models (GLM)

$\bar{x} \rightsquigarrow$ $c = \bar{w}^T \bar{x}$ $\qquad \mu = g^{-1}(c) = \boxed{g^{-1}(\bar{w}^T \bar{x})}$

$c = g(\mu)$

↖ link function

$p(y|\bar{x}, \bar{w})$ $\qquad \mathbb{E}[t(\bar{x})] = g^{-1}(\bar{w}^T \bar{x})$

Overdispersed exponential family

$$p(y|\theta, \sigma^2) = h(y, \sigma^2) \cdot e^{\frac{y\theta - a(\theta)}{\sigma^2}}$$

$= a(\theta)$

$= t(y)$

$$\mathbb{E}\left[\frac{y}{\sigma^2}\right] = \frac{\partial\left(\frac{1}{\sigma^2} a(\theta)\right)}{\partial\theta} \Rightarrow \mathbb{E}[y] = \frac{\partial a(\theta)}{\partial\theta}$$

$$\mathrm{Var}\left[\frac{y}{\sigma^2}\right] = \frac{\partial^2\left(\frac{1}{\sigma^2} a(\theta)\right)}{\partial\theta} \Rightarrow \mathrm{Var}[y] = \sigma^2 \cdot \frac{\partial^2 a(\theta)}{\partial\theta}$$

$\frac{1}{\sigma^4}\mathrm{Var}[y]$

$\theta = \bar{w}^T \bar{x}$

$$g^{-1}(\bar{w}^T \bar{x}) = \mu = \mathbb{E}[y] = a'(\bar{w}^T \bar{x})$$

$$g(\mu) = \bar{w}^T \bar{x}$$

Lin. regr: $g = id$

$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} = e^{\frac{1}{\sigma^2}(y\mu - \frac{1}{2}\mu^2) - \left(\frac{y^2}{2\sigma^2} + \frac{1}{2}\log 2\pi\sigma^2\right)}$$

$$\theta = \mu = \bar{w}^T\bar{x}, \quad a(\theta) = \frac{1}{2}\mu^2, \quad h(y,\sigma^2) = e^{-\left(\frac{y}{2\sigma^2} + \frac{1}{2}\log 2\pi\sigma^2\right)}$$

Logistic regr.

$$p(y|\mu) = \mu^y(1-\mu)^{1-y} = e^{y\log\mu + (1-y)\log(1-\mu)} =$$

$$= e^{y\underbrace{\log\frac{\mu}{1-\mu}} + \log(1-\mu)}, \quad \log(1-\mu) = -a(\theta)$$

$$\sigma^2 = 1, \quad \theta = \log\frac{\mu}{1-\mu} = g(\mu) \qquad e^\theta = \frac{\mu}{1-\mu}, \quad \mu = \frac{e^\theta}{1+e^\theta} =$$

$$a(\theta) = -\log(1-\mu), \quad h(y,\sigma^2) = 1 \qquad = \frac{1}{1+e^{-\theta}}$$

$$p(y|\bar{x},\bar{w},\sigma^2) = h(y,\sigma^2)\, e^{\frac{y(\bar{w}^T\bar{x}) - a(\bar{w}^T\bar{x})}{\sigma^2}}$$

$$D = \{y_1, \dots, y_N\}$$

$$\log p(y_1 \dots y_N | \bar{x}_1 \dots \bar{x}_N, \bar{w}, \sigma^2) = \sum_{n=1}^N \left(\overbrace{\log h(y_n,\sigma^2)}^{\text{const}} + \frac{1}{\sigma^2}\left(y_n \bar{w}^T\bar{x}_n - a(\bar{w}^T\bar{x}_n)\right)\right)$$

$$= \text{const} + \frac{1}{\sigma^2}\cdot\sum_{n=1}^N \left(y_n\cdot\bar{w}^T\bar{x}_n - a(\bar{w}^T\bar{x}_n)\right) \xrightarrow[\bar{w}]{} \max$$

$$\nabla_{\bar{w}} \ell(\bar{w}) = \frac{1}{\sigma^2}\sum_n \left(y_n\bar{x}_n - a'(\bar{w}^T\bar{x}_n)\cdot\bar{x}_n\right) =$$

true prediction

$$= \frac{1}{\sigma^2}\sum_n \left(y_n - \boxed{a'(\bar{w}^T\bar{x}_n)}\right)\bar{x}_n = \frac{1}{\sigma^2}\sum_{n=1}^N (y_n - \mu_n)\,\bar{x}_n$$

$$\boxed{a'(\bar{w}^T\bar{x}_n)} = \mu_n$$

$$H(\bar{w}) = -\frac{1}{\sigma^2}\sum_{n=1}^N \frac{\partial\mu_n}{\partial\theta_n}\cdot\bar{x}_n\bar{x}_n^T = -\frac{1}{\sigma^2}\cdot X^T\cdot\begin{pmatrix} \ddots & & 0 \\ & \frac{\partial\mu_n}{\partial\theta_n} & \\ 0 & & \ddots \end{pmatrix}\cdot X$$

$$H(\bar{w}) = -\frac{1}{\sigma^2} X^T S X$$

$\mu_n(1-\mu_n)$

log. perp.

IRLS

$$\bar{w} := \bar{w} + H^{-1} X^T (\bar{\mu} - \bar{y}) =$$

$$= \bar{w}^{old} - \frac{1}{\sigma^2} (X^T S X)^{-1} X^T (\bar{\mu} - \bar{y}) =$$

$$= (X^T S X)^{-1} X^T S \bar{z}, \quad$$

$$\bar{z} = X\bar{w}^{old} - \frac{1}{\sigma^2} S^{-1} (\bar{\mu} - \bar{y})$$

Poisson regression

$$y \sim Pois(\boxed{\lambda}) = \frac{1}{y!} \lambda^y e^{-\lambda}$$

$\overset{?}{=} \bar{w}^T \bar{x}$

count data

$\mu = \lambda$

$$p(y|\mu) = \frac{1}{y!} \mu^y e^{-\mu} =$$

$$= e^{y \boxed{\log \mu} - \underbrace{\mu}_{``a(\theta)"} - \log y!} \qquad \theta = \log \mu$$

$$\sigma^2 = 1, \quad a(\theta) = e^{\theta} = e^{\bar{w}^T \bar{x}}, \quad h(y, \sigma^2) = \frac{1}{y!}$$

$$\log \mu = \bar{w}^T \bar{x} \qquad \log y_n \approx \bar{w}^T \bar{x}_n$$

Negative binomial

$$NB(k | n, p) \propto p^n (1-p)^k$$

$$p(\theta | D) = \frac{p(\theta) p(D|\theta)}{p(D)}$$