1. Jupyter Notebook PA1
   some questions found in the homework
2 Some review and supplement for lecture

① $\boxed{\text{Exponential family}}$ eg: Gaussian, Bernoulli, Poisson

• $$P(y;\eta) = \underline{b(y)} \exp(\eta^T \underline{T(y)} - a(\eta))$$

η: parameter (natural)

$a(\eta)$: log partition function

y: random variable

$b(y)$: a function of y

$T(y)$: $\boxed{\text{sufficient statistic}}$

"function of samples"

"Contain all information"

归一化因子 (Normalization factor)

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

$$= \frac{1}{\exp(a(\eta))} b(y)\exp(\eta^T T(y))$$

$$\triangleq \frac{1}{Z(\eta)} b(y)\exp(\eta^T T(y))$$

$Z(\eta)$ → partition function

$$\exp(a(\eta)) = Z(\eta)$$

$$a(\eta) = \log Z(\eta)$$

eg: Gaussian distribution: $y_1, y_2, \ldots y_N$ samples

$$T(y) = \begin{pmatrix} \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i^2 \end{pmatrix} \longrightarrow (\mu, \sigma$$

"sufficient"
作用: 压缩数据

• Gaussian $N(\mu, \sigma)$

$$P(y;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \qquad \theta = (\mu, \sigma^2) \qquad \eta = \eta(\theta)$$

$(\mu, \sigma^2 \longrightarrow \eta)$

$$P(y;\eta) = \underline{b(y)} \exp(\eta^T \underline{T(y)} - a(\eta))$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y^2 - 2\mu y + \mu^2)\right\}$$

$$= \exp\log(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\underbrace{(y^2 - 2\mu y)}_{(-2\mu,\ 1)\begin{pmatrix} y \\ y^2 \end{pmatrix}} - \frac{\mu^2}{2\sigma^2}\right\}$$

$$= \exp\log(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(-2\mu \quad 1)\begin{pmatrix} y \\ y^2 \end{pmatrix} - \frac{\mu^2}{2\sigma^2}\right\}$$

$$= \exp\left\{\underbrace{\left(\frac{\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2}\right)}\begin{pmatrix} y \\ y^2 \end{pmatrix} - \underbrace{\left(\frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log 2\pi\sigma^2\right)}\right\}$$

$b(y) = 1$ $\quad \eta^T \quad (T(y)) \quad a(\eta)$

Note: $\begin{cases} N(\mu, 1): & T(y) = y \\ N(\mu, \sigma^2): & T(y) = \begin{pmatrix} y \\ y^2 \end{pmatrix} \end{cases}$ "sufficient"

(GLM)

② $\boxed{\text{Generalized Linear Model}}$

$\begin{cases} \text{i } & y|x;\theta \sim \text{Exponential Family with parameter } \eta \\ \text{ii } & h_\theta(x) = E[T(y)|x;\theta] \\ \text{iii. } & \eta = \theta^T x \end{cases}$

eg: (Linear regression on GLM)

response variable y Gaussian

link function: $\eta = g(\mu) = \mu$

prediction $h_\theta(x) = E[y|x;\theta] = \underbrace{\mu = g^{-1}(\eta)}_{\text{response function}} = \theta^T x$

KEY: ① $y \sim$ exponential family (eg. Gaussian, Bernoulli ......)

② change to exponential family

find relationship between natural parameter (η) and parameter in the concrete distribution (θ)

$\begin{cases} \text{link function}: & \eta = g(\theta) \\ \text{response function}: & \theta = g^{-1}(\eta) \end{cases}$

③ Calculate $E[T(y)|x;\theta]$ and use η to replace (response function)

④ $\eta = \theta^T x$ (GLM)

③ Matrix basic

- trace

$$A=\begin{bmatrix} a_{11} & \cdots & a_{1n}\\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn}\end{bmatrix} \qquad tr(A) = a_{11}+a_{22}+\cdots+a_{nn} = \sum_{i=1}^{n} a_{ii}$$

$$tr(AB) = tr(BA)$$

$$\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij}b_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ji}a_{ij}$$

- $x=\begin{bmatrix} 1 \\ -2 \\ 3\end{bmatrix}$

$$\|x\|_2^2 = 1^2+(-2)^2+3^2 = 14$$

$$\|x\|_1 = |1|+|-2|+|3| = 6$$

$$\|x\|_\infty = 3$$

$$A=\begin{bmatrix} 1 & 2 \\ 3 & 4\end{bmatrix} \qquad \|A\|_F^2 = 1^2+2^2+3^2+4^2 = 30.$$

- $$\frac{\partial \vec{x}}{\partial a} = \begin{bmatrix} \frac{\partial x_1}{\partial a} \\ \vdots \\ \frac{\partial x_n}{\partial a}\end{bmatrix}, \quad \frac{\partial a}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial a}{\partial x_1} \\ \vdots \\ \frac{\partial a}{\partial x_n}\end{bmatrix}, \quad \left[\frac{\partial \vec{x}}{\partial \vec{y}}\right]_{ij} = \left[\frac{\partial x_i}{\partial y_j}\right]$$

- $$\frac{\partial w^T x}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial w^T x}{\partial x_1} \\ \vdots \\ \frac{\partial w^T x}{\partial x_n}\end{bmatrix} = \begin{bmatrix} \frac{\partial (\sum_{i=1}^{n} w_i x_i)}{\partial x_1} \\ \vdots \\ \frac{\partial (\sum_{i=1}^{n} w_i x_i)}{\partial x_n}\end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n\end{bmatrix} = w$$

- $$\frac{\partial x^T A x}{\partial x} = \begin{bmatrix} \frac{\partial x^T A x}{\partial x_1} \\ \vdots \\ \frac{\partial x^T A x}{\partial x_n}\end{bmatrix} = \begin{bmatrix} \frac{\partial \sum_{i=1}^{n}\sum_{j=1}^{n} x_i \cdot a_{ij} \cdot x_j}{\partial x_1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} \cdot x_j + \sum_{i=1}^{n} a_{i1} x_i \\ \vdots \end{bmatrix}$$

$$= (A+A^T)x$$

- $\frac{\partial A^{-1}}{\partial x}$..

$$A^{-1}A = I$$

$$\frac{\partial A^{-1}}{\partial x}\cdot A + A^{-1}\cdot\frac{\partial A}{\partial x} = 0 \qquad \frac{\partial A^{-1}}{\partial x} = -A^{-1}\cdot\frac{\partial A}{\partial x}\cdot A^{-1}$$