Machine Learning: Homework 4

Sen Wang

March 18th, 2019

1 Exponential Family

A. Identify the relevant components necessary for use in a GLM.

a). Normal distribution
Expanding expression for normal distribution, we have,

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \tag{1}$$
$$= \frac{1}{\sqrt{2\pi}}\exp\{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2} - \frac{1}{2\sigma^2}\mu^2 - \ln\sigma\} \tag{2}$$

From above we can obtain relevant components,

$$\mathbf{y} = \{x, x^2\}^T \tag{3}$$
$$\boldsymbol{\theta} = \{\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\}^T \tag{4}$$
$$h(x) = \frac{1}{\sqrt{2\pi}} \tag{5}$$
$$A(\theta) = -\frac{1}{2}\ln(-2\theta_2 - \frac{\theta_1^2}{2\theta_2} \tag{6}$$

b). Binomial distribution
Expanding expression for binomial distribution, we have,

$$p(x|\mu,N) = \binom{N}{x}\mu^x(1-\mu)^{N-x} \tag{7}$$
$$= \binom{N}{x}\exp\{x\ln(\frac{\mu}{1-\mu} + N\ln(1-\mu)\} \tag{8}$$

The relevant components are,

$$y = x \tag{9}$$
$$\theta = \ln(\frac{\mu}{1-\mu}) \tag{10}$$
$$h(x) = \binom{N}{x} \tag{11}$$
$$A(\theta) = N\ln(1+e^\theta) \tag{12}$$

c). Poisson distribution Same as above,

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{13}$$
$$= \frac{1}{x!}\exp\{x\ln\lambda - \lambda\} \tag{14}$$

The relevant components are,

$$y = x \tag{15}$$
$$\theta = \ln \lambda \tag{16}$$
$$h(x) = \frac{1}{x!} \tag{17}$$
$$A(\theta) = \lambda = e^{\theta} \tag{18}$$

d). Gamma distribution Expanding the expression,

$$f(y) = \frac{1}{\Gamma(\nu)}(\frac{\nu}{\mu})^{\nu} y^{\nu-1} e^{-y\nu/\mu} \tag{19}$$
$$= \frac{1}{\Gamma(\nu)}(\frac{\nu}{\mu})^{\nu} \exp\{(\nu - 1)\ln y - \frac{\nu}{\mu} y\} \tag{20}$$

Relevant components are,

$$\mathbf{y} = \{\ln y, y\}^T \tag{21}$$
$$\boldsymbol{\theta} = \{\nu - 1, -\frac{\nu}{\mu}\}^T \tag{22}$$
$$h(y) = 1 \tag{23}$$
$$g(\theta) = \frac{1}{\Gamma(\theta_1 + 1)}(-\theta_2)^{\theta_1 + 1} \tag{24}$$

e). Inverse Gaussian distribution Expanding the expression,

$$f(x) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\} \tag{25}$$
$$= (2\pi x^3)^{-0.5} \exp\{-\frac{\lambda}{2\mu^2} x - \frac{\lambda}{2}\frac{1}{x} + \frac{\lambda}{\mu} + \frac{1}{2}\ln \lambda\} \tag{26}$$

So the relevant components are,

$$\mathbf{y} = \{x, \frac{1}{x}\}^T \tag{27}$$
$$\boldsymbol{\theta} = \{-\frac{\lambda}{2\mu^2}, -\frac{\lambda}{2}\} \tag{28}$$
$$h(x) = (2\pi x^3)^{-0.5} \tag{29}$$
$$A(\theta) = -4\theta_1\theta_2 - \frac{1}{2}\ln 2\theta_2 \tag{30}$$

B. Exponential distribution.
a). Rearranging the expression for exponential distribution,

$$f(y) = \lambda \exp(\lambda y) \tag{31}$$
$$= \exp(\lambda y + \ln \lambda) \tag{32}$$

So the relevant components for GLM are,

$$\theta = \lambda \tag{33}$$
$$h(y) = 1 \tag{34}$$
$$A(\theta) = -\ln \lambda = -\ln \theta \tag{35}$$

b). The canonical link is

$$f(x) = \Psi^{-1}(x) = A'(x) = -\frac{1}{x} \tag{36}$$

c). There is a singularity when $x = 0$ or for regression $\mathbf{w}^T\boldsymbol{\phi} = 0$.

C. Conway-Maxwell Poisson distribution
a). Expanding the expression,

$$P(Y = y) = \frac{\lambda^y}{(y!)^\nu}\frac{1}{z(\lambda, \nu)} \tag{37}$$

$$= \frac{1}{z(\lambda, \nu)} \exp\{y \ln \lambda - \nu \ln(y!)\} \tag{38}$$

So we have the parameters,

$$\mathbf{y} = \{y, \ln(y!)\}^T \tag{39}$$

$$\boldsymbol{\theta} = \{\ln \lambda, -\nu\} \tag{40}$$

$$h(y) = 1 \tag{41}$$

$$g(\boldsymbol{\theta}) = \frac{1}{z(\lambda, \nu)} = (\sum_{i=1}^{\infty} \frac{e^{i\theta_1}}{(i!)^{-\theta_2}})^{-1} \tag{42}$$

$$\tag{43}$$

b). For poisson distribution, there is only one parameter so it can only be fitted with the mean, not the disperison. For CMP Poisson, there are now two parameters therefore both mean and variance can be fitted.

2 Generalized linear model - Probit regression

A. Compute an expression for the posterior. Based on the provided likelihood and the priors, we can easily obtain the following expression by integrating out the σ^2 term,

$$\pi(\mathbf{w}|\mathbf{y},\mathbf{X}) \propto |\mathbf{X}^T\mathbf{X}|^{1/2}\Gamma((2k-1)/4)(\mathbf{w}^T(\mathbf{X}^T\mathbf{X})\mathbf{w})^{-(2k-1)/4}\pi^{-k/2} \tag{44}$$

$$\prod_{i=1}^{n}\Phi(\mathbf{x}^{iT}\mathbf{w})^{y_i}[1-\Phi(\mathbf{x}^{iT})]^{1-y_i} \tag{45}$$

B. Posterior distribution from Metropolis-Hasting algorithm

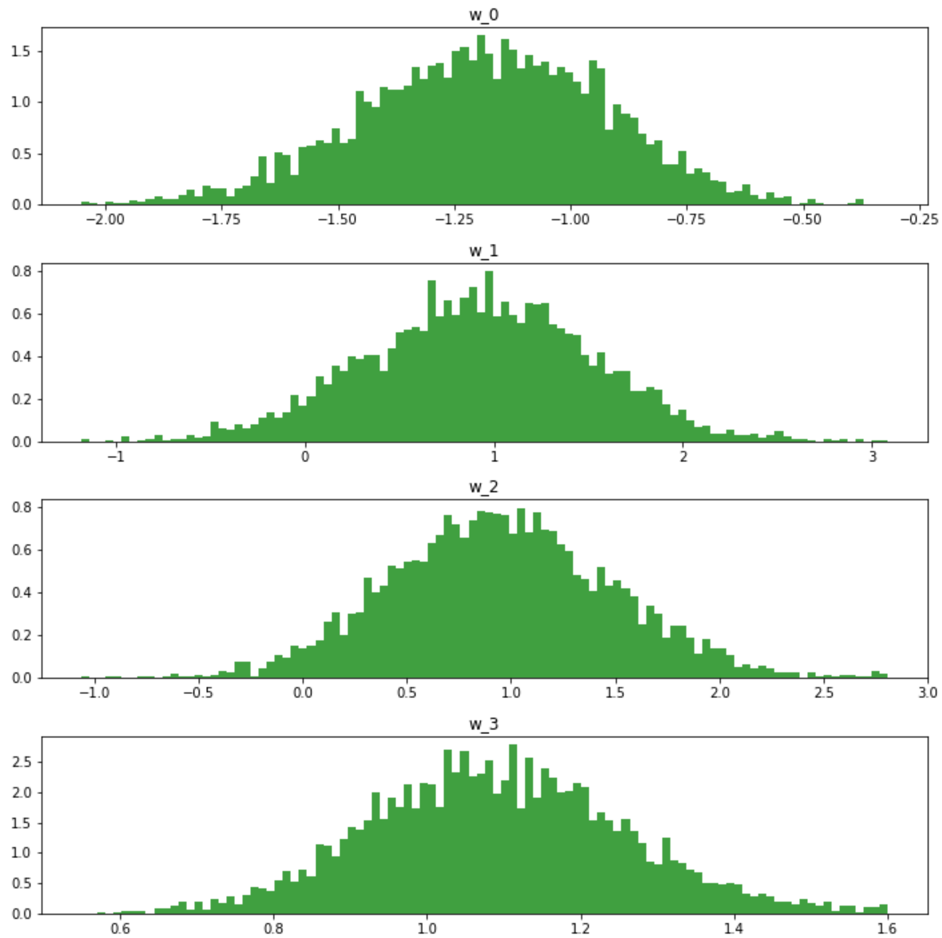

Figure 1: Posterior distribution of $\mathbf{w} from MCMC$

C. By drawing samples from both the posterior distribution of $\mathbf{w}$ and randomly generated $\mathbf{x}$, we are able to produce the predictive distribution as shown below in the diagram,
D. The marginal distribution is obtained through an importance approximation to the marginal probability as shown below,

$$p(\mathbf{y}|\mathbf{X}) \propto \frac{\Gamma(\frac{2k-1}{4})|\mathbf{X}^T\mathbf{X}|^{1/2}}{\pi^{k/2}M}\sum_{m=1}^{M}(\mathbf{w}^{(m)T}(\mathbf{X}^T\mathbf{X})\mathbf{w}^{(m)})^{-\frac{2k+1}{4}}\prod_{i=1}^{n}\Phi(\mathbf{x}^{iT}\mathbf{w}^{(m)})^{y_i}[1-\Phi(\mathbf{w}^{iT}\mathbf{w}^{(m)}]^{1-y_i} \tag{46}$$

$$|\hat{\mathbf{V}}|^{1/2}(4\pi)^{k/2}\exp[(\mathbf{w}^{(m)}-\hat{\mathbf{w}})^T\hat{V}^{-1}\mathbf{w}^{(m)}-\hat{\mathbf{w}})/4] \tag{47}$$

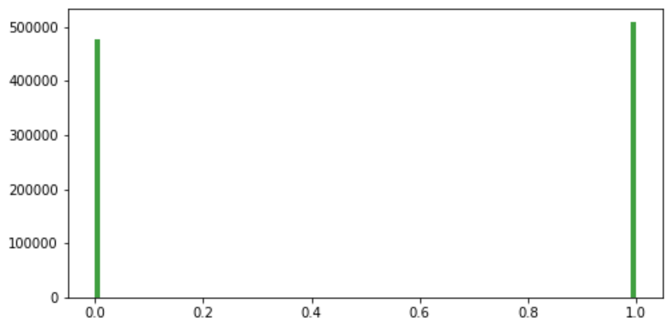

Figure 2: Posterior distribution of $\mathbf{w} from MCMC$

where $\mathbf{w}^{(m)}$ are sampled from the $\mathbf{N}(\hat{\mathbf{w}}, 2\hat{V})$ importance distribution. The detailed implementation please refer to ipython notebook. The Bayes Factor is obtained from the ratio of the marginal probability of $\mathbf{y}$ of the importance sampling and the null hypothesis, $\ln B_{10} = 40$

3 Generalized linear model - Logit regression

A. Compute expression for the posterior Based on the provided likelihood and the priors, we can easily obtain the following expression by integrating out the σ^2 term,

$$\pi(\mathbf{w}|\mathbf{y}, \mathbf{X}) \propto |\mathbf{X}^T\mathbf{X}|^{1/2}\Gamma((2k-1)/4)(\mathbf{w}^T(\mathbf{X}^T\mathbf{X})\mathbf{w})^{-(2k-1)/4}\pi^{-k/2} \tag{48}$$

$$\frac{exp\{\sum_{i=1} y_i \mathbf{x}^{iT}\mathbf{w}\}}{\pi_{i=1}^n[1+\exp(\mathbf{x}^{iT}\mathbf{w})]} \tag{49}$$

Posterior distribution from Metropolis-Hasting algorithm

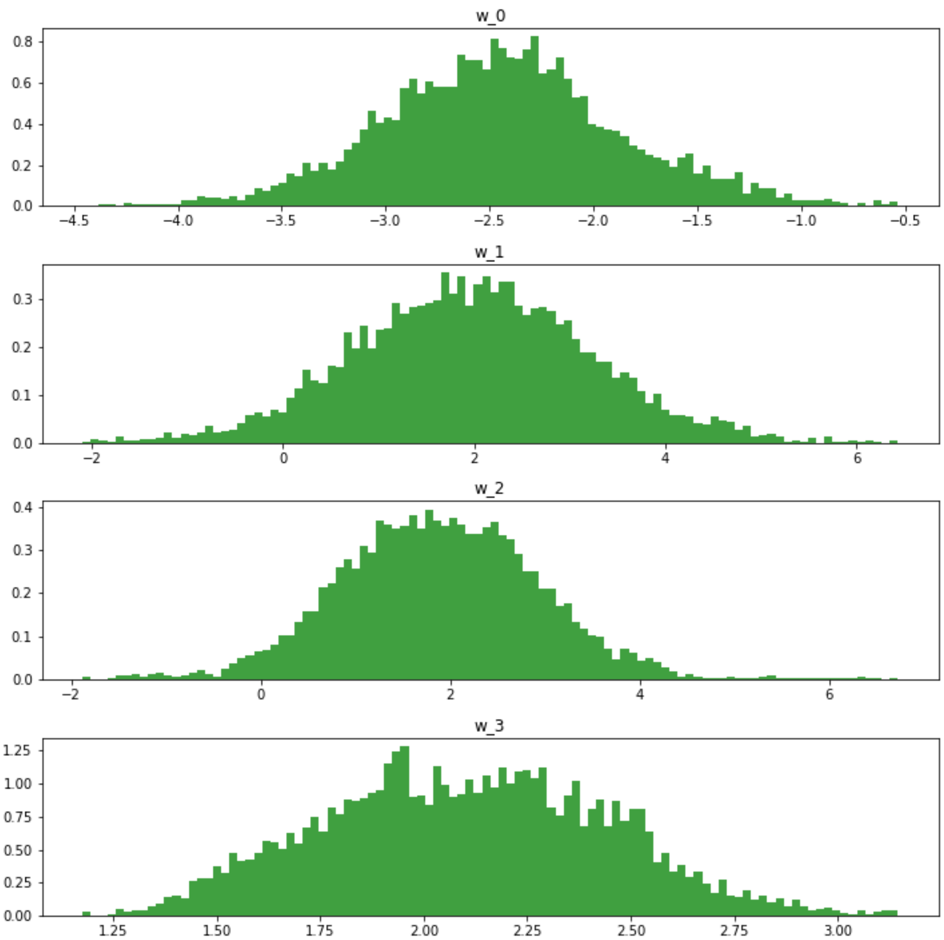

Figure 3: Posterior distribution of $\mathbf{w} from MCMC$

C. By drawing samples from both the posterior distribution of $\mathbf{w}$ and randomly generated $\mathbf{x}$, we are able to produce the predictive distribution as shown below in the diagram,

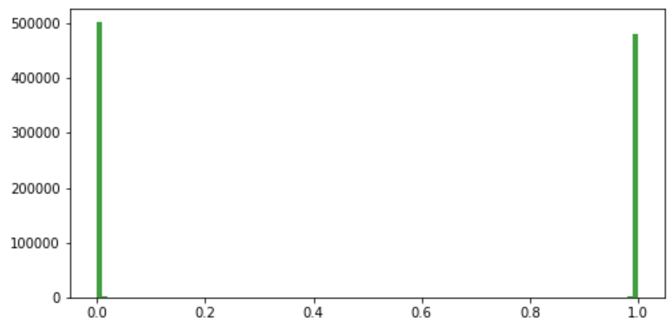

Figure 4: Posterior distribution of $\mathbf{w} from MCMC$

4 K-means algorithm

The K-means algorithm, as detailed in notes pg 11, is implemented in ipython notebook and the resulted cluster means are printed as output.

5 Gaussian mixture, expectation maximization and mixture of experts

A. Show the student's t distribution can be represented as infinite mixture We start with obtaining the expression for the Normal-Gamma distribution,

$$\mathcal{N}(\mu, \frac{sigma^2}{z}) = \frac{1}{Z_N} z \exp[-\frac{z}{2}\frac{x-\mu)^2}{\sigma^2}] \tag{50}$$

$$\mathrm{Ga}(z|\frac{nu}{2}, \frac{nu}{2}) = \frac{Z_G}{z}^{\frac{nu}{2}-1} \exp[-\frac{\nu}{2}z] \tag{51}$$

$$\mathcal{N}(\mu, \frac{sigma^2}{z})\mathrm{Ga}(z|\frac{nu}{2}, \frac{nu}{2}) = \frac{1}{Z} z^{\nu/2} \exp[-u(x)z] \tag{52}$$

$$u(x) = \frac{1}{2}[\frac{(x-\mu)^2}{\sigma^2} + \nu] \tag{53}$$

Integrate with respect to z, we have,

$$\int_0^\infty \frac{1}{Z} z^{\nu/2} \exp[-u(x)z]dz = \frac{1}{Z}\int_0^\infty z^{\nu/2} \exp[-u(x)z]dz \tag{54}$$

Having $t = uz$, we then have,

$$\int_0^\infty (\frac{t}{u})^{\nu/2} \exp(-t)\frac{1}{u}dt = u(x)^{-(1+\frac{\nu}{2})} \int_0^\infty t^{\frac{nu}{2}} \exp(-t)dt \tag{55}$$

$$= u(x)^{-(1+\frac{\nu}{2})}\Gamma(\frac{\nu}{2}+1) \tag{56}$$

Therefore, we obtain the total expression,

$$\frac{1}{Z}\Gamma(\frac{\nu}{2}+1)u(x)^{-(1+\frac{\nu}{2})} = \frac{1}{Z}\Gamma(\frac{\nu}{2}+1)\frac{1}{2}[\frac{x-\mu)^2}{\sigma^2} + \nu]^{-(1+\frac{\nu}{2})} \tag{57}$$

$$= \frac{1}{Zt}[1 + \frac{1}{\nu}\frac{(x-\mu)^2}{\sigma^2}]^{-(1+\frac{\nu}{2})} \tag{58}$$

The term in bracket is the functional form of the Student's t distribution.