Bayesian Models for Machine Learning
Problem Set #1

Si Kai Lee sl3950@columbia.edu

September 30, 2016

Problem 1

Let X_i be the event where door i is chosen, D_i be the event where the prize is behind door i and Y_i be the event where door i is opened by the host. Assume $X = X_1$ and $Y = Y_3$.

$$\begin{aligned}
P(D_1|X_1, Y_3) &= \frac{P(D_1, X_1, Y_3)}{P(X_1, Y_3)} \\
&= \frac{P(D_1, X_1, Y_3)}{P(D_1, X_1, Y_3) + P(D_2, X_1, Y_3)} \\
&= \frac{1/3 * 1/2}{1/3 * 1/2 + 1/3 * 1} \text{as the host cannot open the selected door } P(Y_3|X_1, D_2) = 1 \\
&= 1/3
\end{aligned}$$

Since $P(D_3|X_1, Y_3) = 0$ as the host opened the door to show that it is empty, so $P(D_2|X_1, Y_3) = 1 - 1/3 = 2/3$. Hence the friend should switch.

Problem 2

The conjugate prior is the Dirichlet distribution which is defined as $\frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1}$.

Likelihood: $X_i \sim Mult(\pi)$, Prior: $\pi \sim Dir(\alpha)$. Assume $\forall a_i = \alpha$.

$$\begin{aligned}
p(\pi|X) &= \prod_{j=1}^N \frac{\Gamma(\sum_i x_i + 1)}{\prod_i \Gamma(x_i + 1)} \prod_{i=1}^K \pi_i^{x_j} \cdot \frac{\Gamma\left(\sum_K \alpha\right)}{\prod_K \Gamma(\alpha)} \prod_{i=1}^K \pi_i^{\alpha - 1} \\
&\propto \prod_{j=1}^N \prod_{j=1}^K \pi_i^{x_j} \cdot \prod_{i=1}^K \pi_i^{\alpha-1} = \prod_{i=1}^K \pi_i^{\alpha + n_i - 1} \text{ where } n_i \text{ is the \# of observations in the ith category} \\
&= \frac{\Gamma\left(\sum_{i=1}^K \alpha + n_i\right)}{\prod_{i=1}^K \Gamma(\alpha + n_i)} \prod_{i=1}^K \pi_i^{\alpha + n_i - 1} = Dir(\alpha + n)
\end{aligned}$$

The parameter α_i for the ith class is shifted positively by the number of observations n_i in the same category.

Problem 3

a

We note that $p(\mu, \lambda|X) \propto p(X|\mu, \lambda)p(\mu|\lambda)p(\lambda|b, c)$. Expand $p(x|\mu, \lambda)$ and remove non-μ, λ terms:

$$p(X|\mu, \lambda) = \prod_{i=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left(-\frac{\lambda}{2}(x_i - \mu)^2\right) \propto \lambda^{N/2} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{N}(x_i - \mu)^2\right)$$

Next, expand $p(\mu, \lambda)$ and remove non-μ, λ terms:

$$\begin{aligned} p(\mu, \lambda) = p(\mu|\lambda)p(\lambda|b, c) &= \left(\frac{\lambda}{2\pi a}\right)^{1/2} \exp\left(-\frac{\lambda}{2a}\mu^2\right) \cdot \frac{c^b}{\Gamma(b)}\lambda^{b-1}\exp(-c\lambda) \\ &\propto \lambda^{b-1/2}\exp\left(-\frac{\lambda}{2a}\mu^2 - c\lambda\right) \end{aligned}$$

Due to conjugacy, we know $p(\mu, \lambda|x)$ is a Normal Gamma distribution:

$$\begin{aligned} p(\mu, \lambda|X) &\propto \lambda^{N/2} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{N} x_i^2 + \lambda\mu\sum_{i=1}^{N} x_i - \frac{N\lambda\mu^2}{2}\right) \cdot \lambda^{b-1/2}\exp\left(-\frac{\lambda}{2a}\mu^2 - c\lambda\right) \\ &= N(\mu_n, \lambda_n)\ Gamma(\alpha_n, \beta_n) \end{aligned}$$

Integrating w.r.t. to λ i.e. removing all λ terms, we obtain the distribution of μ:

$$\begin{aligned} p(\mu|\lambda, X) &\propto \exp\left(\lambda\mu\sum_{i=1}^{N} x_i - \frac{N\lambda\mu^2}{2} - \frac{\lambda}{2a}\mu^2\right) \\ &\propto \exp\left[-\frac{\lambda}{2a}\left((aN+1)\mu^2 - 2aN\bar{x}\mu\right)\right] \text{ where } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \\ &\propto \exp\left[-\frac{\lambda}{2a}\left((aN+1)\mu^2 - 2aN\bar{x}\mu + \frac{(aN\bar{x})^2}{aN+1}\right)\right] \\ &\propto \exp\left[-\frac{\lambda(aN+1)}{2a}\left(\mu^2 - 2\frac{aN\bar{x}}{aN+1}\mu + \frac{aN\bar{x}}{aN+1}^2\right)\right] \\ &\propto \lambda^{1/2}\exp\left[-\frac{\lambda_n}{2}(\mu - \mu_n)^2\right] \text{ where } \mu_n = \frac{aN\bar{x}}{aN+1}, \lambda_n = \frac{\lambda}{a}(aN+1) \\ &= N(\mu_n, \lambda_n) \end{aligned}$$

Comparing terms of $p(\mu, \lambda|X)$ and $p(\mu|\lambda, X)$, we find $p(\lambda|X)$:

$$\begin{aligned} p(\lambda|X) &\propto \lambda^{N/2+b-1}\exp\left(-\lambda\left(\frac{1}{2}\sum_{i=1}^{N} x_i^2 + c - \frac{1}{2}\frac{(aN\bar{x})^2}{a(aN+1)}\right)\right) \\ &= Gamma(\alpha_n, \beta_n) \text{ where } \alpha_n = b + N/2,\ \beta_n = c + \frac{1}{2}\sum_{i=1}^{N} x_i^2 - \frac{1}{2}\frac{(aN\bar{x})^2}{a(aN+1)} \end{aligned}$$

b

Let $\kappa_n = \frac{1}{a}(aN+1)$.

$$
\begin{aligned}
&p(x^*|x_{1:n}) \\
&= \int_0^\infty \int_{-\infty}^\infty p(x^*|\mu,\lambda)p(\mu,\lambda|x_{1:n})d\mu d\lambda \\
&= \int_0^\infty \int_{-\infty}^\infty \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left[-\frac{\lambda}{2}(x^*-\mu)^2\right] \left(\frac{\kappa_n\lambda}{2\pi}\right)^{1/2} \exp\left[-\frac{\kappa_n\lambda}{2}(\mu-\mu_n)^2\right] \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\lambda^{\alpha_n-1}\exp[-\beta_n\lambda]d\mu d\lambda \\
&= \int_0^\infty \frac{\kappa_n\lambda^{\alpha_n}}{2\pi}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\exp\left[-\beta_n\lambda-\frac{\lambda}{2}x^{*2}-\frac{\lambda}{2}\mu_n^2\right]\int_{-\infty}^\infty \exp\left[-\frac{\lambda}{2}((d+1)\mu^2-2(x^*+\kappa_n\mu_n)\mu)\right]d\mu d\lambda \\
&= \int_0^\infty Z\int_{-\infty}^\infty \exp\left[-\frac{\lambda(\kappa_n+1)}{2}\left(\mu-\frac{x^*+\kappa_n\mu_n}{\kappa_n+1}\right)^2+\frac{\lambda}{2}\frac{(x^*+\kappa_n\mu_n)^2}{\kappa_n+1}\right]d\mu d\lambda \text{ where } Z \text{ is constant w.r.t. } \mu \\
&= \int_0^\infty Z\exp\left[\frac{\lambda}{2}\frac{(x^*+\kappa_n\mu_n)^2}{\kappa_n+1}\right]\int_{-\infty}^\infty \exp\left[-\left(\mu-\frac{x^*+\kappa_n\mu_n}{\kappa_n+1}\right)^2\Big/\frac{2}{\lambda(\kappa_n+1)}\right]d\mu d\lambda \\
&= \int_0^\infty Z\exp\left[\frac{\lambda}{2}\frac{(x^*+\kappa_n\mu_n)^2}{\kappa_n+1}\right]\left(\frac{2\pi}{(\kappa_n+1)\lambda}\right)^{1/2}d\lambda \\
&= \int_0^\infty \left(\frac{\kappa_n}{\kappa_n+1}\right)^{1/2}(2\pi)^{-1/2}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\lambda^{\alpha_n-1/2}\exp\left[-\beta_n\lambda-\frac{\lambda}{2}x^{*2}-\frac{\lambda}{2}\mu_n^2+\frac{\lambda}{2}\frac{(x^*+\kappa_n\mu_n)^2}{\kappa_n+1}\right]d\lambda \\
&= \int_0^\infty \left(\frac{\kappa_n}{\kappa_n+1}\right)^{1/2}(2\pi)^{-1/2}\frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\lambda^{\alpha_n-1/2}\exp\left[\frac{\lambda}{2(\kappa_n+1)}(\kappa_n x^{*2}+\kappa_n\mu_n^2+2\beta_n(\kappa_n+1)-2\kappa_n\mu_n x^*)\right]d\lambda \\
&= Z'\frac{\Gamma(\alpha_{n+1})}{\beta_{n+1}^{\alpha_{n+1}}}\int_0^\infty \frac{\beta_{n+1}^{\alpha_{n+1}}}{\Gamma(\alpha_{n+1})}\lambda^{\alpha_{n+1}}\exp(-\beta_{n+1}\lambda)d\lambda \text{ where } Z' \text{ is constant w.r.t. } \lambda \text{ and see below for } \alpha_{n+1},\beta_{n+1} \\
&= \left(\frac{\kappa_n}{\kappa_n+1}\right)^{1/2}(2\pi)^{-1/2}\frac{\Gamma(\alpha_{n+1})}{\Gamma(\alpha_n)}\frac{\beta_n^{\alpha_n}}{\beta_{n+1}^{\alpha_{n+1}}}
\end{aligned}
$$

We know that the above corresponds to a t-distribution according to Murphy[1] and

$$
\begin{aligned}
\alpha_{n+1} &= \alpha_n + 1/2 \\
\beta_{n+1} &= \beta_n + \frac{\kappa_n}{2(\kappa_n+1)}(x^*-\mu_n)^2
\end{aligned}
$$

[1] Murphy, Kevin P. "Conjugate Bayesian analysis of the Gaussian distribution." def 1, no. 2?2 (2007): 16.