CS 194-10, Fall 2011
Assignment 5 Solutions

1. *Conjugate Priors (30)*

   (a) Exponential and Gamma
   The likelihood is $P(\mathbf{X} \mid \lambda) = \prod_{i=1}^{N} \lambda \exp(-\lambda x_i)$ and the prior is $p(\lambda \mid \alpha, \beta) = gamma(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{(\alpha-1)} \exp(-\beta\lambda)$. Let $\mathbf{X}$ denote the observations $x_1, ...x_N$ and let s_N denote their sum. Then the posterior is

$$\begin{aligned} p(\lambda \mid \mathbf{X}) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{(\alpha-1)} \exp(-\beta\lambda) \prod_{i=1}^{N} \lambda \exp(-\lambda x_i) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha+N-1} \exp\left(-\lambda(\beta + s_N)\right) \\ &\propto gamma(\lambda \mid \alpha + N, \beta + s_N). \end{aligned}$$

   Therefore the parameter updates are as follows:

$$\begin{aligned} \alpha' &\leftarrow \alpha + N \\ \beta' &\leftarrow \beta + s_N \end{aligned}$$

   For the prediction distribution we compute the following integral:

$$\begin{aligned} p(x_{N+1} \mid x_1, ..., x_N) &= \int p(x_{N+1} \mid \lambda) p(\lambda \mid x_1, ..., x_N) d\lambda \\ &= \int \lambda \exp(-\lambda x_{N+1}) gamma(\lambda \mid \alpha + N, \beta + s_N) d\lambda \\ &= \frac{(\beta + s_N)^{\alpha+N}}{\Gamma(\alpha + N)} \frac{\Gamma(\alpha + N)}{(\beta + s_N + x_{N+1})^{\alpha+N}} \int \lambda gamma(\lambda \mid \alpha + N, \beta + s_N + x_{N+1}) d\lambda \\ &= \frac{(\beta + s_N)^{\alpha+N}}{(\beta + s_N + x_{N+1})^{\alpha+N}} \frac{\alpha + N}{\beta + s_N + x_{N+1}} \end{aligned}$$

   where the penultimate step uses the standard formula α/β for the expected value of a gamma distribution.

   (b) Geometric and Beta
   The likelihood for a single observation of value k is $P(X = k \mid \theta) = (1 - \theta)^{k-1}\theta$ and prior is $p(\theta \mid a, b) = Beta(a, b) = \alpha\theta^{a-1}(1 - \theta)^{b-1}$, where α is the normalization constant. Then the posterior is,

$$\begin{aligned} p(\theta \mid X) &= \alpha\theta^{a-1}(1 - \theta)^{b-1}(1 - \theta)^{k-1}\theta \\ &= \alpha\theta^{a}(1 - \theta)^{b+k-2} \\ &= Beta(\theta \mid a + 1, b + k - 1) \end{aligned}$$

   Therefore the parameter updates are ,

$$\begin{aligned} a' &\leftarrow a + 1 \\ b' &\leftarrow b + k - 1 \end{aligned}$$

For the prediction distribution we compute the following integral:

$$
\begin{aligned}
p(X_2 = \ell \mid X_1 = k) &= \int p(X_2 = \ell \mid \theta) p(\theta \mid X_1 = k) d\theta \\
&= \int (1-\theta)^{\ell-1} \theta Beta(\theta \mid a+1, b+k-1) d\theta \\
&= \frac{\Gamma(a+b+k)}{\Gamma(a+1)\Gamma(b+k-1)} \frac{\Gamma(a+1)\Gamma(b+k+\ell-2)}{\Gamma(a+b+k+\ell-1)} \int \theta Beta(\theta \mid a+1, b+k+\ell-2) d\theta \\
&= \frac{\Gamma(a+b+k)}{\Gamma(b+k-1)} \frac{\Gamma(b+k+\ell-2)}{\Gamma(a+b+k+\ell-1)} \frac{a+1}{a+b+k+\ell-1} \\
&= \frac{\Gamma(a+b+k)}{\Gamma(b+k-1)} \frac{\Gamma(b+k+\ell-2)}{\Gamma(a+b+k+\ell)} \cdot (a+1)
\end{aligned}
$$

where the penultimate step uses the standard formula $\alpha/(\alpha+\beta)$ for the expected value of a Beta distribution.

(c) Mixture Prior

The prior is given by the mixture,

$$
P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_M) = \sum_{m=1}^{M} w_m P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_m) \quad .
$$

Moreover, we are given that $P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_m)$ is a conjugate prior for the likelihood $P(\mathbf{X} \mid \boldsymbol{\theta})$; in other words,

$$
P(\boldsymbol{\theta} \mid \mathbf{X}, \boldsymbol{\gamma}_m) = \alpha_m P(\mathbf{X} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_m) = P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}'_m) \ .
$$

When we multiply the mixture prior with the likelihood, we get the following posterior:

$$
\begin{aligned}
P(\boldsymbol{\theta} \mid \mathbf{X}, \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_M) &= \alpha P(\mathbf{X} \mid \boldsymbol{\theta}) \sum_{m=1}^{M} w_m P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_m) \\
&= \sum_{m=1}^{M} \alpha w_m P(\mathbf{X} \mid \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}_m) \\
&= \sum_{m=1}^{M} \frac{\alpha w_m}{\alpha_m} P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}'_m) \\
&= \sum_{m=1}^{M} w' P(\boldsymbol{\theta} \mid \boldsymbol{\gamma}'_m) \ .
\end{aligned}
$$

Therefore we observe that the posterior has the same form as the prior, i.e., a mixture distribution with updated weights and hyperparameters.

2. Bayesian Naive Bayes

(a) Using the results in 1(a), given a single e-mail sample with attribute values x_j and class y, we perform the following update:

$$
\begin{aligned}
\alpha_{j,y} &\leftarrow \alpha + 1 \\
\beta_{j,y} &\leftarrow \beta + x_j
\end{aligned}
$$