EECS E6720: Bayesian Models for Machine Learning
Homework 1

Josh Rutta

September 24, 2018

Problem 1 My suggestion would be to switch doors.

Without loss of generality, suppose you pick door 1, and the host opens door 2
Call event A_1 = prize behind door 1, and event B_2 = host opening door 2.
$p(A_1) = \frac{1}{3}$ because there are three doors which have equal probability of containing the prize.
According to Bayes' Rule, the probability of getting the prize without switching doors would be

$$p(A_1|B_2) = \frac{p(B_2|A_1)p(A_1)}{p(B_2)}$$

$p(B_2|A_1) = \frac{1}{2}$, because if the prize was behind door 1, the host could have equally chosen door 2 or door 3.
$p(B_2) = \sum_{i=1}^{3} p(B_2, A_i) = \sum_{i=1}^{3} p(B_2|A_i)p(A_i) = \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2}$

($p(B_2|A_2) = 0$ because the host won't open door 2 if the prize is behind it,, and $p(B_2|A_3) = 1$ because if you initially pick door 1 and prize is behind door 3, the host must open door 2)

$\Rightarrow$p(A$_1|B_2) = \frac{\frac{1}{2}\times\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$.
The probability of getting the prize with switching would be

$$p(A_3|B_2) = \frac{p(B_2|A_3)p(A_3)}{p(B_2)}$$

The only probability different here is $p(B_2|A_3) = 1$ because, as explained before, if you initially pick door 1 and prize is behind door 3, the host must open door 2 ($p(A_3) = p(A_1)$ and $p(B_2)$ remains the same).
$\Rightarrow$p(A$_3|B_2) = \frac{1\times\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

Problem 2

2. $\pi=(\pi_1,\dots,\pi_K)$, $\pi_j \geq 0$, $\Sigma_j \pi_j = 1$. $X_i \sim \text{Multinomial}(\pi)$, i.i.d. for $i=1,\dots,N$

$$p(\pi \mid x_1, X_2, \dots, X_N) \propto p(X_1,\dots,X_N \mid \pi)\, p(\pi)$$

$$p(X_1,\dots,X_N \mid \pi) = p(X_1 \mid \pi)\, p(X_2 \mid \pi) \cdots p(X_N \mid \pi)$$ because i.i.d.

$$p(X_i \mid \pi) = \frac{k!}{\prod_{i=1}^{K} x_i!} \prod_{i=1}^{K} \pi_i^{x_i}$$

$$\Rightarrow p(\pi \mid X_1,\dots,X_N) \propto \prod_{i=1}^{N} \left(\prod_{j=1}^{K} \pi_j^{x_{ij}} \right) p(\pi)$$

$$\propto \left(\prod_{j=1}^{K} \pi_j^{x_{1j}} \right) \left(\prod_{j=1}^{K} \pi_j^{x_{2j}} \right) \cdots \left(\prod_{j=1}^{K} \pi_j^{x_{Nj}} \right) p(\pi)$$

$$p(\pi \mid X_1,\dots,X_N) \propto \left(\prod_{j=1}^{K} \pi_j^{\sum_{i=1}^{N} x_{ij}} \right) p(\pi)$$

Since posterior contains product of exponentials, conjugate prior should be Dirichlet

$$p(\pi) = \frac{1}{B(\vec{\alpha})} \prod_{j=1}^{K} \pi_j^{\alpha_j - 1}$$

$$\Rightarrow p(\pi \mid X_1,\dots,X_N) \propto \prod_{j=1}^{K} \pi_j^{\sum_{i=1}^{N} x_{ij} + \alpha_j - 1}$$

$$\alpha_j' = \sum_{i=1}^{N} x_{ij} + \alpha_j$$

$$\Rightarrow p(\pi \mid X_1,\dots,X_N) = \frac{1}{B(\vec{\alpha}_j')} \prod_{j=1}^{K} \pi_j^{\alpha_j' - 1}$$

normalizes distribution

product of exponentials is most obvious feature of distribution

Problem 3

data model

3. a) $x_i \in \mathbb{N}$, $x_i \overset{iid}{\sim} \text{Poisson}(\lambda)$

model prior : $\lambda \sim \text{Gamma}(a,b)$

$$p(\lambda \mid x_1,\dots,x_N) \propto p(x_1,\dots,x_N \mid \lambda)\, p(\lambda)$$

$$p(x_1,\dots,x_N \mid \lambda) = \prod_{i=1}^{N} \text{poisson}(\lambda)$$

$$= \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$p(x_1,\dots,x_N \mid \lambda) = \frac{e^{-\lambda N}\lambda^{\sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} x_i!}$$

$$p(\lambda) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)}$$

$$\Rightarrow p(\lambda \mid x_1,\dots,x_N) \propto \left(e^{-\lambda N}\lambda^{\sum_{i=1}^{N} x_i}\right)\left(\lambda^{a-1} e^{-b\lambda}\right)$$

$$p(\lambda \mid x_1,\dots,x_N) \propto e^{-\lambda(N+b)}\lambda^{\sum_{i=1}^{N} x_i + a - 1}$$

$$a' = \sum_{i=1}^{N} x_i + a$$

$$b' = N + b$$

$$p(\lambda \mid x_1,\dots,x_N) \propto e^{-\lambda b'}\lambda^{a'-1}$$

To normalize, $p(\lambda \mid x_1,\dots,x_N) = \dfrac{b'^{a'}\lambda^{a'-1}e^{-\lambda b'}}{\Gamma(a')}$

$$= \text{Gamma}(a', b')$$

$$p(\lambda \mid x_1,\dots,x_N) = \text{Gamma}\left(\sum_{i=1}^{N} x_i + a,\ b+N\right)$$

b) $p(x^* \mid x_1,\dots,x_n) = \int_0^\infty p(x^* \mid \lambda)\, p(\lambda \mid x_1,\dots,x_N)\, d\lambda$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^{x^*}}{x^*!} \cdot \frac{(b+N)^{\sum_{i=1}^{N} x_i + a}\, \lambda^{\sum_{i=1}^{N} x_i + a - 1}\, e^{-\lambda(b+N)}}{\Gamma\left(\sum_{i=1}^{N} x_i + a\right)}\, d\lambda$$

$$= \frac{(b+N)^{\sum_{i=1}^{N} x_i + a}}{x^*!\,\Gamma\left(\sum_{i=1}^{N} x_i + a\right)} \int_0^\infty e^{-\lambda(b+N+1)}\lambda^{\sum_{i=1}^{N} x_i + x^* + a - 1}\, d\lambda \, ; \quad \int_0^\infty \text{Gamma}(a,b)\, d\lambda = \int_0^\infty \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)}\, d\lambda \ 1$$

$$\Rightarrow \int_0^\infty \lambda^{a-1} e^{-b\lambda}\, d\lambda = \frac{\Gamma(a)}{b^a}$$

$$= \frac{(b+N)^{\sum_{i=1}^{N} x_i + a}\ \Gamma\left(\sum_{i=1}^{N} x_i + x^* + a\right)}{x^*!\,\Gamma\left(\sum_{i=1}^{N} x_i + a\right)(b+N+1)^{\sum_{i=1}^{N} x_i + x^* + a}}$$

Problem 4 (a and b)
(work to derive posterior)

In 3) $p(x_1,\dots,x_N|\lambda) = \dfrac{e^{-\lambda N}\lambda^{\sum_{i=1}^N x_i}}{\prod_{i=1}^N x_i!}$ $\qquad \dfrac{(a+b)!}{a!}$ $\qquad n! = n(n-1)!$ $\quad \dfrac{(a+b)(a+b-1)\dots \cancel{a!}}{\cancel{a!}}$

$$p(\lambda) = \frac{b^a\lambda^{a-1}e^{-b\lambda}}{\Gamma(a)}$$

$$p(\lambda|x_1,\dots,x_N) = Gamma\left(\sum_{i=1}^N x_i + a,\ b+N\right)$$

$$p(x^*|x_1,\dots,x_N) = \frac{(b+N)^{\sum_{i=1}^N x_i + a}\,\Gamma\left(\sum_{i=1}^N x_i + x^* + a\right)}{x^*!\,\Gamma\left(\sum_{i=1}^N x_i + a\right)(b+N+1)^{\sum_{i=1}^N x_i + x^* + a}}$$

In 4. $p(x_n|\vec{\lambda}_1, y_n=1) = \prod_{d=1}^{54} \dfrac{e^{-\lambda_{1,d}}\lambda_{1,d}^{x_{n,d}}}{x_{n,d}!}$

$$p(\lambda_{1,d}) \overset{iid}{=} Gamma(1,1) = \frac{b^a\lambda^{a-1}e^{-b\lambda}}{\Gamma(a)} = e^{-\lambda_{1,d}}$$

$\pi = Beta(1,1) = 1$ on $[0,1]$

$$p(y^*=y|x^*,X,\vec{y}) \propto p(x^*|y^*=y,\{x_i:y_i=y\})\,p(y^*=y|\vec{y})$$

$$p(x^*|y^*=y,\{x_i:y_i=y\}) = \prod_{d=1}^{54}\int_0^\infty p(x^*|\lambda_{y,d})\,p(\lambda_{y,d}|\{x_i:y_i=y\})\,d\lambda$$

fixed d for $y=1$: $\displaystyle\int_0^\infty p(x^*|\lambda_{1,d})\,p(\lambda_{1,d}|\{x_i:y_i=1\})\,d\lambda = \int_0^\infty \frac{e^{-\lambda_{1,d}}\lambda_{1,d}^{x^*_d}}{x^*_d!}\, Gamma\left(\sum_{i:y_i=1} x_{i,d}+1,\ 1+N_1\right)d\lambda$ $\qquad N_1 = (\# y==1)$

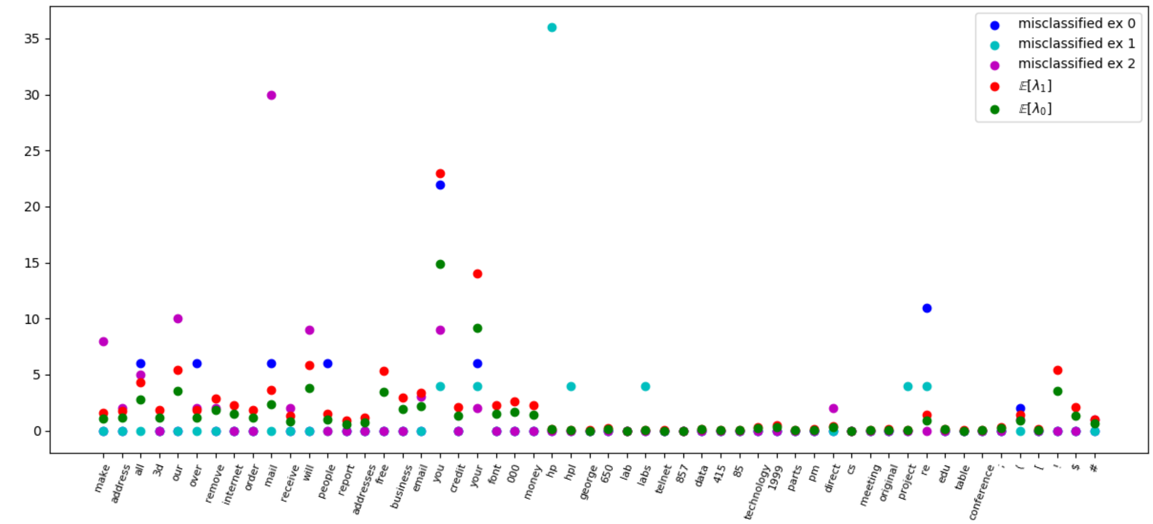
acc. to 3, this equals: $\displaystyle\frac{(1+N_1)^{\sum_{i:y_i=1} x_{i,d}+1}\,\Gamma\left(\sum_{i:y_i=1} x_{i,d} + x^*_d + 1\right)}{x^*_d!\,\Gamma\left(\sum_{i:y_i=1} x_{i,d}+1\right)(N_1+2)^{\sum_{i:y_i=1} x_{i,d}+x^*_d+1}} = \xi_d$

$$p(x^*|y^*=y,\{x_i:y_i=1\}) = \left(\prod_{d=1}^{54}\xi_d\right)$$

$$\mathbb{E}[\lambda_{y,d}] = \frac{\sum_{i:y_i=y} x_{i,d}+1}{1+N_y}$$

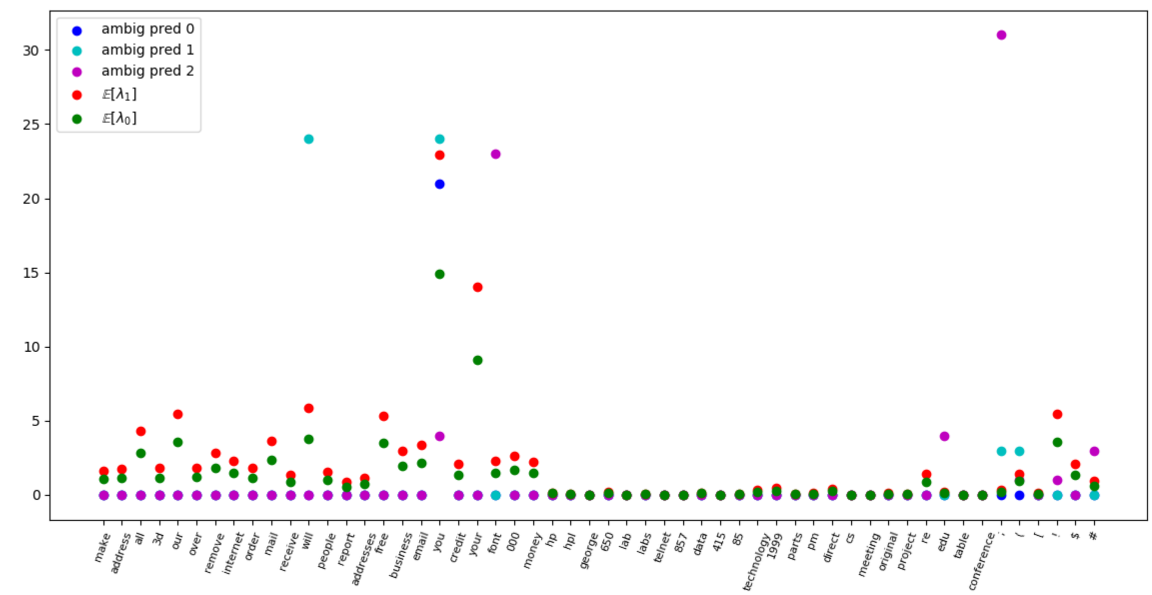
	non-spam	spam
classified as non-spam	231	11
classified as spam	48	171

(c)



misclassified example 0 : $p(y^* = 1|x^*, X, y) = 0.9999999205074594, p(y^* = 0|x^*, X, y) = 7.949254068798455e - 08$
misclassified example 1 : $p(y^* = 1|x^*, X, y) = 1.1315568628386464e - 84, p(y^* = 0|x^*, X, y) = 1.0$
misclassified example 2 : $p(y^* = 1|x^*, X, y) = 1.0, p(y* = 0|x*, X, y) = 1.7065245215354857e - 23$

(d)



ambiguous pred 0 : $p(y^* = 1|x^*, X, y) = 0.4893797226855785, p(y^* = 0|x^*, X, y) = 0.5106202773144215$
ambiguous pred 1 :$p(y^* = 1|x^*, X, y) = 0.514217610001227, p(y^* = 0|x^*, X, y) = 0.4857823899987729$
ambiguous pred 2 :$p(y^* = 1|x^*, X, y) = 0.47692775612020744, p(y^* = 0|x^*, X, y) = 0.5230722438797926$