Solutions by **Andrew Lys** `andrewlys (at) u.e.`

1. Gaussian Mixtures

2. Modeling Text Documents

2.1. A Simple Model.

(a) We shall denote p_{topic} as p, since it is given that this is a single probability. For simplicity, we assume that $y \in \{0, 1\}$, and that $x \in \{0, 1\}^N$. We denote $x[i]$ to be the ith coordinate of the sample x.

Given a sample

$$S = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

We define the following sample statistics. For $x \in \{0, 1\}$, $y \in \{0, 1\}$:

$$n_j(y, x) = |\{i : (x_i, y_i) \in S, x_i[j] = x, y_i = y\}|$$
$$n(y) = |\{i : (x_i, y_i) \in S, y_i = y\}|$$

We want to find estimators for p and for

$$P(x[1] = x_1, \dots, x[N] = x_N | y = y)$$

By the independence of $x[i]|y$, we can simplify this expression:

$$P(x[1] = x_1, \dots, x[N] = x_N | y = y) = \prod_{i=1}^{N} P(x[i] = x_i | y = y)$$

Thus, we can focus on estimators of p and $P(x[i] = x | y = y) := p_i(x|y)$.

We should expect that our MLEs for p and $p_i(x|y)$ to be the sample means, i.e.

$$\hat{p} = \frac{n(1)}{n}$$
$$\hat{p}_i(x|y) = \frac{n_i(y, x)}{n(y)}$$

We define our log-likelihood function as

$$\ell(\theta|S) = \sum_{i=1}^{n} \log(P(y = y_i, x[1] = x_i[1], \dots, x[N] = x_i[N]))$$

Given that S was drawn i.i.d., we can simplify.

$$\begin{aligned}
\ell(\theta|S) &= \sum_{i=1}^{n} \log(P(x[1] = x_i[1], \dots, x[N] = x_i[N] | y = y_i) P(y = y_i)) \\
&= \sum_{i=1}^{n} \log(P(y = y_i) \prod_{j=1}^{N} P(x[j] = x_i[j] | y = y_i)) \\
&= \sum_{i=1}^{n} \log(P(y = y_i)) + \sum_{j=1}^{N} \log(P(x[j] = x_i[j] | y = y_i)) \\
&= \sum_{i=1}^{n} \log(P(y = y_i)) + \sum_{i=1}^{n} \sum_{j=1}^{N} \log(p_j(x_i[j] | y_i))
\end{aligned}$$

Writing the parameters explicitly, we have:

$$\ell(\theta|S) = \sum_{i=1}^{n} \log(P(y = y_i | p)) + \sum_{i=1}^{n} \sum_{j=1}^{N} \log(P(x[i] = x_j[i] | y_i, p_i(x|y)))$$

To solve for the minmimum of $\ell(\theta|S)$, we use the method of Lagrange multipliers. First, we can split the problem into two steps. It's clear that that right sum does not depend on p, so we can begin by finding the optimal p.

We note:

$$\begin{aligned}
P(y = y_i|p) &= P(y = y_i|y_i = 1, p)P(y_i = 1|p) + P(y = y_i|y_i = 0, p)P(y_i = 0|p) \\
&= P(y = 1|p)[[y_i = 1]] + P(y = 0|p)[[y_i = 0]] \\
&= p^{y_i}(1-p)^{1-y_i}
\end{aligned}$$

Plugging this into our log-likelihood, we have:

$$\begin{aligned}
\ell(\theta|S) &= \sum_{i=1}^{n} \log(p^{y_i}(1-p)^{1-y_i}) + \sum_{i=1}^{n}\sum_{j=1}^{N} \log(P(x[i] = x_j[i]|y_i, p_i(x|y))) \\
&= \sum_{i=1}^{n} y_i \log(p) + (1-y_i)\log(1-p) + \sum_{i=1}^{n}\sum_{j=1}^{N} \log(P(x[i] = x_j[i]|y_i, p_i(x|y)))
\end{aligned}$$

Taking the derivative with respect to p and setting it to zero, we have:

$$\begin{aligned}
\frac{d}{dp}\ell(\theta|S) &= \sum_{i=1}^{n} \frac{y_i}{p} - \frac{1-y_i}{1-p} = 0 \\
\sum_{i=1}^{n} \frac{y_i}{p} &= \sum_{i=1}^{n} \frac{1-y_i}{1-p} \\
\frac{1-p}{p} &= \frac{\sum_{i=1}^{n} 1-y_i}{\sum_{i=1}^{n} y_i} \\
p &= \frac{\sum_{i=1}^{n} y_i}{n}
\end{aligned}$$

Thus, we have that $\hat{p} = \frac{n(1)}{n}$.

Now, we solve for $\hat{p}_i(x|y)$, by using the method of Lagrange multipliers. Our objective function is as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{N} \log(P(x[j] = x_i[j]|y_i))$$

We can write this in a nicer form.

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{N} \log(P(x[j] = x_i[j]|y_i)) &= \sum_{i=1}^{n}\sum_{j=1}^{N} \log(p_j(x_i[j]|y_i)) \\
&= \sum_{j=1}^{N}\sum_{i=1}^{n}\sum_{x\in\{0,1\}} [[x_i[j] = x]] \log(p_j(x|y_i)) \\
&= \sum_{j=1}^{N}\sum_{i=1}^{n}\sum_{x\in\{0,1\}}\sum_{y\in\{0,1\}} [[x_i[j] = x \wedge y_i = y]] \log(p_j(x|y)) \\
&= \sum_{j=1}^{N}\sum_{x\in\{0,1\}}\sum_{y\in\{0,1\}} \log(p_j(x|y)) \sum_{i=1}^{n}[[x_i[j] = x \wedge y_i = y]] \\
&= \sum_{j=1}^{N}\sum_{y\in\{0,1\}}\sum_{x\in\{0,1\}} \log(p_j(x|y)) n_j(y, x)
\end{aligned}$$

We now have the following constraints:

$$\sum_{x\in\{0,1\}} p_j(x|y) = 1 \qquad \forall y \in \{0,1\}, j \in [N]$$

This gives us the following Lagrangian:

$$\mathcal{L} = \sum_{j=1}^{N} \sum_{y\in\{0,1\}} \sum_{x\in\{0,1\}} \log(p_j(x|y))n_j(y,x) + \sum_{j=1}^{N} \sum_{y\in\{0,1\}} \lambda_j(y) \left(\sum_{x\in\{0,1\}} p_j(x|y) - 1 \right)$$

Taking the derivatives with respect to $p_j(x|y)$, we have:

$$\begin{aligned}
[p_j(x|y)] &: \frac{n_j(y,x)}{p_j(x|y)} = \lambda_j(y) \\
[\lambda_j(y)] &: \sum_{x\in\{0,1\}} p_j(x|y) = 1
\end{aligned}$$

Since we have equality, in $\lambda_j(y)$, for $x \in \{0,1\}$ we can solve for $p_j(x|y)$:

$$\begin{aligned}
\frac{n_j(y,x)}{p_j(x|y)} &= \frac{n_j(y,1-x)}{p_j(1-x|y)} \\
p_j(1-x|y) &= \frac{n_j(y,1-x)}{n_j(y,x)} p_j(x|y) \\
\implies 1 &= p_j(x|y) + \frac{n_j(y,1-x)}{n_j(y,x)} p_j(x|y) \\
n_j(y,x) &= p_j(x|y)n_j(y,x) + n_j(y,1-x)p_j(x|y) \\
&= p_j(x|y)(n_j(y,x) + n_j(y,1-x)) \\
p_j(x|y) &= \frac{n_j(y,x)}{n_j(y,x) + n_j(y,1-x)} \\
\hat{p}_j(x|y) &= \frac{n_j(y,x)}{n(y)}
\end{aligned}$$

(b) Using Baye's Law, and conditional independence we have:

$$\begin{aligned}
P(Y=1|X=x) &= \frac{P(X=x|Y=1)P(Y=1)}{P(X=x)} \\
&= \frac{P(X[1]=x[1],\ldots,X[N]=x[N]|Y=1)P(Y=1)}{P(X[1]=x[1],\ldots,X[N]=x[n])} \\
&= \frac{P(Y=1)\prod_{i=1}^{N} P(X[i]=x[i]|Y=1)}{P(X[1]=x[1],\ldots,X[N]=x[n]|Y=1)P(Y=1) + P(X[1]=x[1],\ldots,X[n]=x[n]|Y=0)P(Y=0)} \\
&= \frac{p\prod_{i=1}^{N} p_i(x[i]|1)}{p\prod_{i=1}^{N} p_i(x[i]|1) + (1-p)\prod_{i=1}^{N} p_i(x[i]|0)}
\end{aligned}$$

Now we can reduce this into the form of a logistic function.

$$\begin{aligned}
P(Y=1|X=x) &= \frac{p\prod_{i=1}^{N} p_i(x[i]|1)}{p\prod_{i=1}^{N} p_i(x[i]|1) + (1-p)\prod_{i=1}^{N} p_i(x[i]|0)} \\
&= \frac{1}{1 + \frac{1-p}{p}\frac{\prod_{i=1}^{N} p_i(x[i]|0)}{\prod_{i=1}^{N} p_i(x[i]|1)}} \\
&= \frac{1}{1 + e^{-(\log(\frac{p}{1-p}) + \sum_{i=1}^{N} \log(\frac{p_i(x[i]|1)}{p_i(x[i]|0)}))}}
\end{aligned}$$

Therefore, we can get our discriminant as follows:

$$r(x) = \log\left(\frac{p}{1-p}\right) + \sum_{i=1}^{N} \log\left(\frac{p_i(x[i]|1)}{p_i(x[i]|0)}\right)$$