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. Additionally, we denote

$$p_y(x_i) = P(x_i = 1|Y = y)$$

Given a sample

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

Some of the x_i's repeat, so let $\{(x_j, y_j)\}_{j=1}^k$ be the elements of the sample such that each (x_i, y_i) occurs only once. Then let

$$n_{y,j} = |\{i : (x_j, y_j) = (x_i, y_i) \in S\}|$$

And similarly, we define

$$n_y = |\{i : y_i = y, (x_i, y_i) \in S\}|$$

Then, we should expect that our MLEs for p and $\{p_y\}$ to be the sample errors, i.e.

$$\hat{p} = \frac{n_1}{n}$$
$$\hat{p}_y(x_i) = \frac{n_{y,i}}{n_y}$$

We derive this with the MLEs estimators.

$$L(p, \{p_y\}|S) = P(S|p, \{p_y\})$$

First, we prove that these samples are independent.

$$\begin{align}
P(x_j, y_j|x_i, y_i) &= P(x_j, y_j|x_i) \tag{1}\\
&= P(x_j|y_j, x_i)P(y_j|x_i) \tag{2}\\
&= P(x_j|y_j)P(y_j) = P(x_j, y_j) \tag{3}
\end{align}$$

(1) is true because y_j is chosen independently of y_i, and x_j does not depend on y_j. (2) is true by the definition of conditional probability. (3) is true because $x_j|y_j$ is conditionally independent of x_i.

Therefore, we have:

$$\begin{aligned}
L(p, \{p_y\}|S) &= \prod_{i=1}^n P(x_i, y_i|p, \{p_y\})\\
&= \prod_{i=1}^n P(x_i, y_i|y_i = 1, p, \{p_y\})P(y_i = 1|p, \{p_y\}) + P(x_i, y_i|y_i = -1, p, \{p_y\})P(y_i = -1|p, \{p_y\})\\
&= \prod_{i=1}^n (P(x_i|y_i = 1, \{p_y\})p)^{\frac{1+y_i}{2}} (P(x_i|y_i = -1, \{p_y\})(1-p))^{\frac{1-y_i}{2}}\\
&= \prod_{i=1}^n (p_1(x_i)p)^{\frac{1+y_i}{2}} (p_{-1}(x_i)p)^{\frac{1-y_i}{2}}
\end{aligned}$$

Taking the log-likelihood, $\ell(p, \{p_y\}|S)$, we have:

$$\ell(p, \{p_y\}|S) = \frac{1}{2}\sum_{i=1}^n (1 + y_i)(\log(p_1(x_i)) + \log(p)) + (1 - y_i)(\log(p_{-1}(x_i)) + \log(1 - p))$$

Taking the derivative, with respect to p, and setting to 0, we get:

$$\begin{aligned}
\frac{\partial}{\partial p}\ell &= \frac{1}{2}\sum_{i=1}^{n}\frac{1+y_i}{p} - \frac{1-y_i}{1-p} = 0 \\
\implies \frac{1}{p}\sum_{i=1}^{n} 1+y_i &= \frac{1}{1-p}\sum_{i=1}^{n} 1-y_i \\
\frac{1-p}{p} &= \frac{n_{-1}}{n_1} \\
\implies p &= \frac{n_1}{n_{-1}+n_1} = \frac{n_1}{n}
\end{aligned}$$

This verifies the first part of our intuition.

For the second part, we consider the following LaGrangian:

$$\mathcal{L} = 2\ell(p, \{p_y\}|S) - \lambda\left(-1 + \sum_{i=1}^{n} p_1(x_i)\right) - \mu\left(-1 + \sum_{i=1}^{n} p_{-1}(x_i)\right)$$

Taking the critical points with respect to $p_y(x_i)$, we get

$$\begin{aligned}
[p_1(x_i)] : \sum_{j:(x_j,y_j)=(x_i,1)} \frac{1+y_j}{p_1(x_j)} &= \lambda \sum_{j:x_j=x_i} p_1(x_j) & \implies \sum_{j:(x_j,y_j)=(x_i,1)} \frac{2}{p_1(x_i)} &= \lambda \sum_{j:x_j=x_i} p_1(x_j) \\
\implies &
\end{aligned}$$