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

$$p_y[i] = P(x_i = 1 | Y = y)$$

Given a sample

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

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.

$$\hat{p} = \frac{n_1}{n}$$
$$\hat{p}_y[i] = \frac{n_{y,i}}{n_y}$$

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.

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

(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