

DS4420: Exercise on Expectation Maximization via Naive Bayes

Recall that Naive Bayes defines the following model. Here, c denotes a class (say, ‘spam’ or ‘not spam’), π comprises class prior probabilities, θ_c are per-word conditional probabilities for class c (assume k classes), and w_n just denotes a particular word drawn from a vocabulary of symbols that we will call \mathcal{V} . We omit the instance index i here.

$$p(c|w_{1:N}, \pi, \theta) \propto p(\pi_c) \prod_{n=1}^N p(w_n|\theta_c) \quad (1)$$

Assume:

$$\begin{aligned} \mathcal{V} &= \{a, b, c\} \\ x_0 &= \{aab\} \\ x_1 &= \{cbc\} \\ x_2 &= \{aaa\} \\ x_3 &= \{cbcc\} \end{aligned}$$

We will assume $k = 2$, i.e., two classes. The task then is to perform *hard* Expectation Maximization to estimate parameters.

Init step. Assume we randomly initialize parameters as follows.

$$\begin{aligned} \pi_0 &= 0.4; \pi_1 = 0.6 \\ p(a|c=0) &= 0.6; p(b|c=0) = 0.3; p(c|c=0) = 0.1 \\ p(a|c=1) &= 0.3; p(b|c=1) = 0.4; p(c|c=1) = 0.3 \end{aligned}$$

E-Step. Update the latent variables (assignments).

M-Step. Update parameter estimates.

Bonus. How would this change for ‘soft’ EM?