Maximum Likelihood.

$$h_{ML} = \underset{h}{\operatorname{argmax}}\, P(d|h)$$

$$P(d|h_{\theta,\theta_1,\theta_2}) = \theta^{c} * (1-\theta)^{l} * \theta_1^{r_c} * (1-\theta_1)^{g_c} * \theta_2^{r_l} * (1-\theta_2)^{g_l}.$$

likelihood of data given the hypothesis.

$$L(d|h_{\theta,\theta_1,\theta_2}) = [c\log\theta + l\log(1-\theta)] + [r_c\log\theta_1 + g_c\log(1-\theta_1)] + [r_l\log\theta_2 + g_l\log(1-\theta_2)].$$

log-likelihood of data given the hypothesis

$$\theta = \underset{\theta}{\operatorname{argmax}}\, L(d|h_{\theta,\theta_1,\theta_2})$$

$$\frac{\partial L(d|h_{\theta,\theta_1,\theta_2})}{\partial\theta} = 0 \Rightarrow \theta = ?$$

$$\frac{\partial L}{\partial\theta} = \frac{c}{\theta} - \frac{l}{1-\theta} = 0$$

$$c(1-\theta) - l\theta = 0$$
$$c - c\theta - l\theta = 0$$
$$\theta(c+l) = c$$
$$\theta = \frac{c}{c+l} = \frac{c}{N}$$

$$\frac{\partial L}{\partial\theta_1} = 0 \Rightarrow \theta_1 = \frac{r_c}{r_c+g_c} = \frac{r_c}{c}$$

$$\frac{\partial L}{\partial\theta_2} = 0 \Rightarrow \theta_2 = \frac{r_l}{r_l+g_l} = \frac{r_l}{l}$$

1 ML parameter learning decomposes into separate learning problems, one for each parameter.

2 Parameter values for a variable, given its parents, are just the observed frequencies of variable values for each setting of parent values.

Zero probability problem:

- When the data set is small enough, that some events have not yet been observed, the ML hypothesis assigns zero probability to those events

solution (Laplace smoothing) initialize the counts for each event to be 1 instead of 0.

complete data:

- each data point contains values for every variable in the probability model.

Learning w/ Hidden Variables : The EM Algorithm.

$$\theta = \frac{\overbrace{\cancel{\text{Observed}}}^{\text{Expected} \rightarrow \text{E-step}} \text{ \# of candies from Bag 1}}{\text{Total \# of candies } (=10)} \searrow \text{M-step.}$$

Expected # of candies from Bag 1

$$= \sum_{j=1}^{10} P(Bag=1 \mid Flavour_j)$$

$$= 7 * P(Bag=1 \mid F=cherry) + 3 * P(Bag=1 \mid F=lime)$$

$$P(Bag=1 \mid F=cherry) = \frac{P(F=cherry \mid Bag=1)\, P(Bag=1)}{\sum_i P(F=cherry \mid Bag=i)\, P(Bag=i)}$$

$$= \frac{\theta\theta_1}{\theta\theta_1 + (1-\theta)\theta_2} = \frac{0.5*0.8}{0.5*0.8 + 0.5*0.3} = 0.73.$$

$$P(Bag=1 \mid F=lime) = 0.22$$

$$\sum_{j=1}^{10} P(Bag=1 \mid Flavour_j) = 7*0.73 + 3*0.22 = 5.77.$$

$$\theta^{(1)} = \frac{5.77}{10} = 0.58$$

$$\theta_1^{(1)} = \frac{\text{Expected \# of cherry candies from Bag 1}}{\text{Expected \# of candies from Bag 1}} = \frac{5.11}{5.77}$$

$$= 0.89.$$

Expected # of cherry candies from Bag 1

$$\sum_{j:F_j=\text{cherry}} P(Bag=1 \mid F_j=\text{cherry}) = 7 * 0.73 = 5.11$$

$$\theta_2^{(1)} = \frac{\text{Expected \# of cherry candies from Bag 2}}{\text{Expected \# of candies from Bag 2}} = \frac{1.89}{4.23} = 0.45.$$

Expected # of cherry candies from Bag 2

$$\sum_{j=1, F_j=\text{cherry}}^{N} P(Bag=2 \mid F_j=\text{cherry}) = 7 * 0.27 = 1.89.$$

Expected # of candies from Bag 2 $= 10 - 5.77 = 4.23$

The EM Algorithm. ML: $\underset{\theta}{\text{argmax}}\, P(d \mid h_\theta)$

$$\theta^{(i+1)} = \underset{\theta}{\arg\max} \sum_{z} P(Z=z \mid x, \theta^{(i)}) L(x, Z=z \mid \theta)$$

↑ M-step ↑ E-step.

x: observed values in all the examples.

Z: all the hidden variables for all the examples

θ: all the parameters for the probability model.

E-step: Compute the expected values of hidden variables for each example.

M-step: Compute the parameters to maximize the expected log likelihood.