CM146, Winter 2018
Problem Set xx: yy

1 Problem 1

(a) **Solution:**
TODO

(b) **Solution:**
We know from Conditional Probability that

$$P(D_i, y_i) = P(D_i|y = 0) \cdot P(y = 0) + P(D_i|y = 1) \cdot P(y = 1)$$

$$P(y_i = 1) = \theta, P(y_i = 0) = 1 - \theta$$

$$P(D_i, y_i) = \left(\frac{n!}{a_i!b_i!c_i!}\alpha_1^{a_i}\beta_1^{b_i}\gamma_1^{c_i}\right) \cdot \theta + \left(\frac{n!}{a_i!b_i!c_i!}\alpha_0^{a_i}\beta_0^{b_i}\gamma_0^{c_i}\right) \cdot (1 - \theta)$$

$$\log P(D_i, y_i) =$$

2 Problem 2

(a) **Solution:**
The two unspecified transition probabilities are:

- $q_{22} = P(q_{t+1} = 2|q_t = 2) = 1 - P(q_{t+1} = 1|q_t = 2) = 1 - 1 = 0$
- $q_{21} = P(q_{t+1} = 2|q_t = 1) = 1 - P(q_{t+1} = 1|q_t = 1) = 1 - 1 = 0$

The two unspecified outcome probabilities are:

- $e_1(B) = P(O_t = B|q_t = 1) = 1 - P(O_t = A|q_t = 1) = 1 - 0.99 = 0.01$
- $e_2(A) = P(O_t = A|q_t = 2) = 1 - P(O_t = B|q_t = 2) = 1 - 0.51 = 0.49$

(b) **Solution:**
The probability that the first symbol is A can be calculated as follows:

$$P(first = A) = P(O_1 = A|q_1 = 1) \cdot P(q_1 = 1) + P(O_1 = A|q_1 = 2) \cdot P(q_1 = 2)$$

$$P(first = A) = (0.99)(0.49) + (0.49)(0.51) = 0.735$$

The probability that the first symbol is B can be calculated as follows:

$$P(first = B) = P(O_1 = B|q_1 = 1) \cdot P(q_1 = 1) + P(O_1 = B|q_1 = 2) \cdot P(q_1 = 2)$$

$$P(first = B) = (0.01)(0.49) + (0.51)(0.51) = 0.265$$

Hence, the more frequent output in the first position is A

(c) **Solution:**
TODO

3 Problem 3

(a) The problem with minimizing over μ, c, k is that we are extremely prone to overfitting. With this kind of loss function, we will generate a model that achieves $J(\mu, c, k) = 0$ on a training set of size n, by using n cluster centers, where each cluster has the label that is assigned to it in the training data. This is an overfitted model and would perform terribly on test sets.
For a training set S with labels y,

$$\mu = S$$

$$c = y$$

$$k = |S|$$