Q.5) Naive Bayes with Binomial features:

a) Maximum likelihood.

$$l(\theta) = \log \prod_{i=1}^{m} p(x^{(i)} \mid y^{(i)}; \theta)\, P(y^{(i)})$$

$$= \log \prod_{i=1}^{m} \left[\prod_{j=1}^{n} P(x_j^{(i)} \mid y^{(i)}; \theta) \right] P(y^{(i)})$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \log P(x_j^{(i)} \mid y^{(i)}; \theta) + \sum_{i=1}^{m} P(y^{(i)})$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \log \binom{p^{(i)}}{x_j^{(i)}} \alpha_{j|y=y^{(i)}}^{x_j^{(i)}} \left(1 - \alpha_{j|y=y^{(i)}}\right)^{p^{(i)} - x_j^{(i)}}$$

$$+ \sum_{i=1}^{m} \log P(y^{(i)})$$

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\; l(\theta)$$

$$\frac{\partial l}{\partial \theta} = 0 \quad \text{i.e.} \quad \frac{\partial l}{\partial \alpha_{j|y=l}} = 0$$

$$\alpha_{j|y=l} = \frac{\sum_{i=1}^{m} 1(y^{(i)} = l)\, x_j^{(i)}}{\sum_{i=1}^{m} 1(y^{(i)} = l)\, p^{(i)}}$$