1. Sequence Prediction:

Complexity:

Question A: For Naive Algorithm, the time complexity is:

$$O(M \cdot L^M)$$

where L is the number of states and M is the number of observations. Therefore, there is L^M possible classes, which each costs M time to calculate.

Question B: For Viterbi Algorithm, the time complexity is:

$$O(M \cdot L^2)$$

where L is the number of states and M observations. In Viterbi Algorithm, we use dynamic programming which saves old ~~com~~ consumptions and uses them to make future calcalation. So at each step we just need to consider $L \times L$ combinations of what the next one could be and choosing the max probability. Then mutiplies by M times observations.

Concepts:

Question C: True, this is because when we increase the number of hidden states, we increase the number of possible state sequence which we can choose from, This will increase our ~~likely~~ likelihood that we can exactly match the training data.

Question D. In EM ~~Mar~~ Algorithm, we update our transition matrix during the maximatization step, When updating, we sum up the marginals for each cell, where the marginals are computed as $\alpha\beta$ (α from forward algorithm) (β from backward algorithm)

(i) If a coefficient of the initial state probability matrix is 0, then this coefficient will remain 0 at the end of the EM algorithm. This is because when in forward algorithm, if the initial value in one row is 0, that means the initial probability of having that state ~~with~~ for any observation is 0, so when we do forward algorithm and sum up the values of previous columns, the emission probability will always remain 0. Thus, ~~we~~ when you update the matrix during maximization step by summing the marginals, nothing happens to coefficients that are initial 0, the marginals for these coefficient is 0.

(ii) When a coefficient for a given state of the state transition matrix is 0, then that coefficient will remain 0 until the end of the EM Algorithm, This is because in forward algorithm, the value of that row in the forward matrix is 0. The $P_{initial}$ (transition to that state) = 0. So when you sum up that transitioning to that state from all the state in previous columns, we get 0. Thus this coeffecient remains 0.

2. Naive Bayes

Question A:

In this problem, we have 3 features: color, type and origin. To determine whether a Red Domestic SUV is more likely to be stolen or not, we should calculate:

$P(Red \mid Yes)$, $P(SUV \mid Yes)$, $P(Domestic \mid Yes)$

$P(Red \mid No)$, $P(SUV \mid No)$, $P(Domestic \mid No)$

① Yes:

Red:	SUV:	Domestic:
$n = 3$	$n = 3$	$n = 3$
$n_c = 3$	$n_c = 0$	$n_c = 2$
$p = 0.5$	$p = 0.5$	$p = 0.5$
$m = 3$	$m = 3$	$m = 3$

② No:

Red:	SUV:	Domestic:
$n = 7$	$n = 7$	$n = 7$
$n_c = 2$	$n_c = 4$	$n_c = 3$
$p = 0.5$	$p = 0.5$	$p = 0.5$
$m = 3$	$m = 3$	$m = 3$

Therefore, we can calculate the possibilities using $P(a_i \mid v_j) = \frac{n_c + mp}{n + m}$

$$P(Red \mid Yes) = \frac{3 + 3 \cdot 0.5}{3 + 3} = 0.75$$

$$P(SUV \mid Yes) = \frac{0 + 3 \cdot 0.5}{3 + 3} = 0.25$$

$$P(Domestic \mid Yes) = \frac{2 + 3 \cdot 0.5}{3 + 3} = \frac{7}{12}$$

$$P(Red \mid No) = \frac{2+3 \cdot 0.5}{7+3} = 0.35$$

$$P(SUV \mid No) = \frac{4+3 \cdot 0.5}{7+3} = 0.55$$

$$P(Domestic \mid No) = \frac{3+3 \cdot 0.5}{7+3} = 0.45$$

Therefore:

①. $P(Stolen) \prod P(a_i \mid v_j)$

$= P(Yes) \cdot P(Red \mid Yes) \cdot P(SUV \mid Yes) \cdot P(Domestic \mid Yes)$

$= 0.3 \cdot 0.75 \cdot 0.25 \cdot \frac{7}{12} = 0.0328$

②. $P(not\ stolen) \prod P(a_i \mid v_j)$

$= P(No) \cdot P(Red \mid No) \cdot P(SUV \mid No) \cdot P(Domestic \mid No)$

$= 0.7 \cdot 0.35 \cdot 0.55 \cdot 0.45 = 0.0606$

As $0.0606 > 0.0326$, this car is more likely not to be stolen.

Moreover, as $V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod (a_i \mid v_j)$

Here V_{nb} is ~~not stolen~~ "No".

Question B:

It does not work well

Because the Conditional Indepedence assumption does not hold.