CS698O: Quiz-5

Name: _____ **Roll No.:** _____

1. For a first order HMM, consider transition probabilities at two different steps: $P(Y_{i+1} = c_j \mid Y_i = c_i)$ and $P(Y_{i+5} = c_j \mid Y_{i+4} = c_i)$.
   Which of the following are (is) true? Fill all that you think are correct.

   A: $P(Y_{i+1} = c_j \mid Y_i = c_i) > P(Y_{i+5} = c_j \mid Y_{i+4} = c_i)$

   B: $P(Y_{i+1} = c_j \mid Y_i = c_i) < P(Y_{i+5} = c_j \mid Y_{i+4} = c_i)$

   C: $P(Y_{i+1} = c_j \mid Y_i = c_i) = P(Y_{i+5} = c_j \mid Y_{i+4} = c_i)$

   D: $P(Y_{i+1} = c_j \mid Y_i = c_i) = 5 * P(Y_{i+5} = c_j \mid Y_{i+4} = c_i)$

---

2. What is the decoding problem in case of HMMs? Fill all that you think are correct.

   A: Give the parameters of the distribution what is the probability of the observed sequence?

   B: Given the observations what is the most likely hidden states that explains the observation?

   C: Given the HMM, what are transition and emission probability distributions?

   D: $\underset{y_i \in \Lambda}{\text{argmax}}\, P(Y_i = y_i \mid X_{1:N})$

---

3. For an HMM of sequence length N and K possible states. Approximately, what is the maximum number of sequences possible? Fill all that you think are correct.

   A: $\sim K^N$

   B: $\sim N^K$

   C: $\sim K * N^2$

   D: $\sim N * K^2$

---

4. Decoding in sequence models like HMM and MEMMs can be made efficient using which of the following techniques? Fill all that you think are correct.

   A: Bayes Rule

   B: Dynamic Programing

   C: Kneser-Ney Smoothing

   D: Forward-Backward Computations

---

5. Given only the observations, which of the following techniques can be used for estimating the hidden states of an HMM. Fill all that you think are correct.

   A: Buam Welch Algorithm

   B: Forward-Backward Algorithm

   C: EM Algorithm

   D: MLE Algorithm

---

6. What is the primary motivation for using Maximum Entropy Markov Models (MEMMs) as compared to HMMs? Fill all that you think are correct.

A: MEMMs allow efficient inference over hidden states

B: MEMMs allow to encode rich features into the model

C: MEMMs make use of maximum entropy assumption, so are true model of the sequence probability

D: MEMMs do not suffer from label bias problem

---

7. Consider the probability distribution of a sequence of observed symbols $X_{1:N}$ with hidden state sequence $Y_{1:N}$.

$P_A(Y_{1:N}) = \prod_{i=1}^{N+1} P(Y_i \mid Y_{i-1}) \times P(X_i \mid Y_i)$

$P_B(Y_{1:N}) = \prod_{i=1}^{N+1} P(Y_i \mid Y_{i-1}, X_{1:N})$

Match above probability distributions with the model type. Fill all that you think are correct.

A: $P_A(Y_{1:N})$: MEMM

B: $P_A(Y_{1:N})$: HMM

C: $P_B(Y_{1:N})$: MEMM

D: $P_B(Y_{1:N})$: HMM

---

8. Which of the following are the reason(s) for label-bias problem? Fill all that you think are correct.

A: Conservation of source probability mass

B: Local normalization

C: Hidden states with low entropy

D: Bias in the emission probabilities

---

9. Which of the following are true about Conditional Random Field (CRF) model? Fill all that you think are correct.

A: CRF suffers from label bias problem

B: CRF does not suffers from label bias problem

C: CRF training does not scale well with data

D: CRF performs local normalization at each hidden state

---

10. HMMs suffer from which of the following problems? Fill all that you think are correct.

A: Overfitting

B: Label Bias Problem

C: Inference algorithm grows exponentially with sequence length

D: Conservation of source probability mass
