

Practice questions for Quiz 7

CS228 - Winter 2009

February 21, 2009

The following questions are provided as examples of the types of questions will ask on the quiz this week, as well as the kinds of questions you should be asking yourself (and answering) as you read/review the text. You are encouraged to collaborate on this practice quiz material.

1 Bayesian Prediction with Dirichlet prior

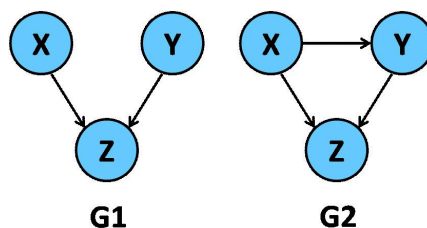
Consider an experiment where we toss a thumbtack multiple times (independently). We model the probability distribution as

$$P_{thumbtack}(X : \theta) = \begin{cases} \theta & \text{if } X = \text{Head} \\ 1 - \theta & \text{otherwise} \end{cases}$$

If we observe 20 heads and 30 tails, then

- What is the MLE of θ ?
- Assume we now add a *Dirichlet*(25, 25) prior, what is the Bayesian prediction of the probability that the outcome of the next toss is head?
- What is the relation between these two estimations? Can the Bayesian prediction with *Dirichlet* prior converge to MLE? If yes, what's the condition(s)?

2 Structure Scores



We want to compare the two graph structures using some scoring functions and given a dataset \mathcal{D} .

- Write down the likelihood scores for these two graphs in terms of the mutual information and entropy in the empirical distribution observed in the dataset \mathcal{D} . What's the difference between them, i.e., $score_L(\mathcal{G}_1 : \mathcal{D}) - score_L(\mathcal{G}_2 : \mathcal{D})$?
- Write down the BIC scores for these two graphs.
- Suppose the dataset \mathcal{D} satisfies all independencies implied by \mathcal{G}_1 , what can you say about the values of

$$* score_L(\mathcal{G}_1 : \mathcal{D}) - score_L(\mathcal{G}_2 : \mathcal{D})$$

$$* score_{BIC}(\mathcal{G}_1 : \mathcal{D}) - score_{BIC}(\mathcal{G}_2 : \mathcal{D})$$

3 Structure Search

Consider performing a structural search over a graph \mathcal{G} , using only three possible operations: adding an edge, deleting an edge, and reversing an edge. Assume the scoring function can be decomposed as the sum of FamScores.

- Performing a single operation changes how many local scores?
- Write down the score changes corresponding to perform each of the three operations.
- What's the criteria to choose an optimal operation in the current search step?
- Can this heuristic greedy search guarantee to find the global maxima given a dataset \mathcal{D} ?

In this quiz, we will also ask a question about BDe prior, so you may want to review it in the 2/19 lecture.