×

Bayesian Priors for BNs

4/4 得分 (100%)

测验通过！

返回第 1 周课程

---

1 / 1 分

1.

BDe Priors. The following is a common approach for defining a parameter prior for a Bayesian network, and is referred to as the BDe prior. Let P_0 be some distribution over possible assignments $x_1, \ldots, x_n$, and select some fixed α . For a node X with parents $\mathbf{U}$ we define $\alpha_{x|\mathbf{u}} = \alpha P_0(x, \mathbf{u})$.

For this question, assume X takes one of m values and that X has k parents, each of which takes d values. If we choose P_0 to be the uniform distribution, then what is the value of $\alpha_{x|\mathbf{u}}$?

- ○ $\alpha/(mk^d)$
- ○ α
- ◉ $\alpha/(md^k)$

正确

For any joint distribution it must hold that $\sum_x \sum_{\mathbf{u}} P_0(x, \mathbf{u}) = \mathbf{1}$; and for a uniform distribution all md^k terms in the sum are constant and hence must equal to $\frac{1}{md^k}$.

- ○ $\alpha/((m+k)d)$

1 / 1 分

2.

Learning with a Dirichlet Prior. Suppose we are interested in estimating the distribution over the English letters. We assume an alphabet that consists of 26 letters and the space symbol, and we ignore all other punctuation and the upper/lower case distinction. We model the distribution over the 27 symbols as a multinomial parametrized by $\theta = (\theta_1, \ldots, \theta_{27})$ where $\sum_i \theta_i = 1$ and all $\theta_i \geq 0$.

Now we go to Stanford's Green library and repeat the following experiment: randomly pick up a book, open a page, pick a spot on the page, and write down the nearest symbol that is in our alphabet. We use $X[m]$ to denote the letter we obtain in the m th experiment.

In the end, we have collected a dataset $D = \{x[1], \ldots, x[2000]\}$ consisting of 2000 symbols, among which "e" appears 260 times. We use a Dirichlet prior over θ , i.e. $P(\theta) = Dirichlet(\alpha_1, \ldots, \alpha_{27})$ where each $\alpha_i = 10$. What is the predictive probability that letter "e" occurs with this prior? (i.e., what is $P(X[2001] = "e" | D)$?) Write your answer as a decimal rounded to the nearest **ten thousandth** (0.xxxx).

0.1189

正确回答

1 / 1 分

3.

Learning with a Dirichlet Prior. In the setting of the previous question, suppose we had collected $M = 2000$ symbols, and the number of times "a" appeared was 100, while the number of times "p" appeared was 87. Now suppose we draw 2 more samples, X[2001] and X[2002]. If we use $\alpha_i = 10$ for all i , what is the probability of $P(X[2001] = "p", X[2002] = "a" | D)$? (round your answer to the nearest **millionth**, 0.xxxxxx)

0.002070

正确回答
Using the chain rule, this breaks down to $P(X[2001] =" p " |D). P(X[2002] =" a " |X[2001] =" p ", D)$. Using this formation and using the updated estimates and total count in $P(X[2002] =" a " |X[2001] =" p ", D)$, we get the correct option as $\frac{97}{2270} \frac{110}{2271}$ which rounds to .002070

---

1 / 1 分

4.
***Learning with a Dirichlet Prior.** In the setting of previous two questions, suppose we have collected M symbols, and let $\alpha = \sum_i \alpha_i$ (we no longer assume that each $\alpha_i = 10$). In which situation(s) does the Bayesian predictive probability using the Dirichlet prior (i.e., $P(X[M+1] \mid D)$) converge to the MLE estimation for any distribution over M ? You may select 1 or more options.

- [ ] Both α and M are fixed and non-zero for some fixed distribution over α

未选择的是正确的

- [x] $\alpha \to 0$ and M is fixed

正确
The Dirichlet prior is a weighted average of the prior mean and the MLE estimate. Thus, if $\alpha \to 0$ for a fixed value of M then the probability will be dominated by the actual counts as the influence of our prior vanishes and will converge to MLE estimation.

- [ ] $M \to 0$ and α is fixed and non-zero

未选择的是正确的

☐ None of the above

未选择的是正确的

☑ $M \to \infty$ and α is fixed

正确

The Dirichlet prior is a weighted average of the prior mean and the MLE estimate. Thus, if $M \to \infty$ for a fixed value of α then the probability will be dominated by the actual counts and will converge to MLE estimation.