Stat Learn & Graph Models

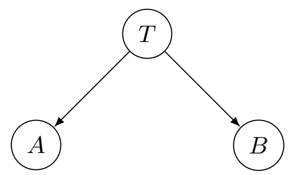
HW 2

Due date: 10/10/2017

Total points: 100 + 5*
UG grading : $\min(p, 100) \times 100/95$
Grad grading : p
where p is the points obtained in the HW.

Problem 1 Maximum likelihood and Bayesian Parameter Estimation for a Bayesian Network

Consider the following Bayesian network:



where each variable takes on either 0 or 1. For example, $T = 0$ or $T = 1$. Furthermore, consider the following parametrization:

$$p(T) = \begin{cases} t, & T = 0 \\ *, & T = 1 \end{cases}$$

$$p(A|T) = \begin{cases} a, & T = 0, A = 0 \\ *, & T = 0, A = 1 \\ a^2 & T = 1, A = 0 \\ * & T = 1, A = 1 \end{cases}$$

$$p(B|T) = \begin{cases} b, & T = 0, B = 0 \\ *, & T = 0, B = 1 \\ b' & T = 1, B = 0 \\ * & T = 1, B = 1 \end{cases}$$

Notation: $a^2 = a \cdot a$ and a $*$ is a placeholder for a value that will make the expression a valid probability distribution.

Suppose we have collected data $\mathcal{D} = \{(T_1, A_1, B_1), \ldots, (T_i, A_i, B_i), \ldots, (T_{100}, A_{100}, B_{100})\}$ from 100 days and the number of times each configuration occurs is given in the following table,

TAB	000	001	010	011	100	101	110	111
# days	34	3	4	2	18	11	12	16

where for example, 001 denotes $T = 0, A = 0, B = 1$.

a) (10 pts) Find the maximum likelihood estimates for b and b'.

b) (5 pts) Find a graphical model that represents the unknown parameters a, b, b', t along with $\mathcal{D}$, and the associated probability distribution.

c) (15 pts) Find the posterior distribution for $p(a|\mathcal{D})$ assuming a Beta(2, 1) prior. You do not need to normalize your answer.

Problem 2 Bayesian Estimation for Normal Prior and Likelihood

Suppose $y_1, \ldots, y_n$ are n independent observations each with distribution $\mathcal{N}\left(\mu, \sigma^2\right)$. Assume $\mu \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$.

a) (15 pts) Prove that

$$p\left(\mu | y_1^n\right) \sim \mathcal{N}\left(b/a, 1/a\right),$$

where

$$b = \frac{\mu_0}{\sigma_0^2} + \frac{\bar{y}}{\sigma^2/n}, \quad (\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i), \qquad a = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}$$

b) (5 pts *) Given the data $y_1, \ldots, y_n$, what is the expected value for the future observation y_{n+1}, that is $E[y_{n+1}|y_1^n]$?

Problem 3 Dirichlet prior and multinomial likelihood

This problem is an extension of the Beta prior and Binomial likelihood. Suppose that we are given a die whose probability of showing i is θ_i for $i = 1, \ldots, 6$. We throw the die n times and i shows n_i times. Our goal is to find the Bayesian estimates of $\theta = (\theta_1, \ldots, \theta_6)$. A common prior for this type of problem is the Dirichlet distribution:

$$(\theta_1, \ldots, \theta_6) \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_6)$$

$$p(\theta_1, \ldots, \theta_6) \propto \prod_{i=1}^{6} \theta_i^{\alpha_i - 1} \qquad \text{for } \theta_i > 0 \text{ and } \sum_{i=1}^{6} \theta_i = 1$$

where the α_i are the parameters of the distribution. More generally, the Dirichlet distribution can have any number of parameters, instead of 6. Note that if instead of 6 parameters, we had two, this would be the same as the Beta distribution: Dirichlet is indeed a generalization of Beta. The mean for each parameter in the above distribution is $E[\theta_i] = \frac{\alpha_i}{\sum_i \alpha_i}$.

a) (10 pts) Prove that the posterior distribution for θ, $p(\theta_1, \ldots, \theta_6 | n_1, \ldots, n_6)$ has a Dirichlet distribution. Find the parameters of this distribution.

b) (10 pts) Choose the values of α_i to represent a non-informative uniform prior. With these values, find the posterior mean for each θ_i.

Problem 4 ML for exponential distribution

Suppose that the lifespan t of a certain type of lamp is given by an exponential distribution with mean θ. That is

$$p(t|\theta) = \frac{1}{\theta} e^{-t/\theta}, \qquad t > 0.$$

Suppose that we have measured the lifespans of n lamps, resulting in independent values $t_1, \ldots, t_n$.

a) (10 pts) Find the ML estimate of θ.

b) (10 pts) Show that this ML estimator is unbiased.

Problem 5 The Exponential Family

Recall that the general form of a distribution from the exponential family for a random variable y and parameter θ is

$$p(y|\theta) \propto \exp\left(a(y)^T b(\theta) + f(y) + g(\theta)\right),$$

where a, b, f, g are given functions and $b(\theta)$ is called the natural parameter. Note that a, b, y, θ can be vectors and f, g are scalars.

Find a, b, f, g for the following distributions:

a) (8 pts) Gaussian distribution, with μ as parameter (and σ as a known constant):

$$p\left(y|\mu\right) \propto \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

b) (7 pts) Gaussian distribution, with μ and σ as parameters:

$$p\left(y|\mu,\sigma^2\right) \propto \frac{1}{\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$