CS5691: PRML Tutorial – 3

Topics: Maximum Likelihood Estimate (MLE), Maximum A Posteriori Estimation (MAP), Bayesian Parameter Estimation

27^{th} Aug 2021

Question 1

What is Maximum Likelihood Estimation?

Question 2

Given N data points from a Gaussian distribution, $\mathcal{D} = \{x_1, ..., x_N\}$ find the following:

Question 2, Part (a)

The Maximum Likelihood Estimate (MLE) mean μ of the distribution.

Question 2, Part (b)

The MLE variance σ^2 of the distribution.

Question 2, Part (c)

Suppose the prior distribution of the mean is also a Gaussian with mean μ_p and variance σ_p^2, find the MAP estimate of the mean.

Question 3

Given the following distributions, mention the parameters we would typically estimate:

a Gaussian Distribution

b Beta Distribution

c Exponential Distribution

d Gamma Distribution

Question 4

Let x have a uniform density

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

Suppose that n samples $\mathcal{D} = \{x_1, ..., x_n\}$ are drawn independently according to $p(x|\theta)$. Show that the MLE for θ is $\max[\mathcal{D}]$, i.e. the value of the maximum element in $\mathcal{D}$.

Question 5

An MLE estimator is a variation of MAP estimator – true or false. Justify your answer.

Question 6

How does a MAP estimator differ from an MLE estimator?

Question 7

Can we use MLE for constrained optimization problem?

Question 8

The purpose of this problem is to derive the Bayesian classifier for the d-dimensional multivariate Bernoulli case. Let the conditional probability for a given category be given by

$$P(\mathrm{x}|\theta) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{(1-x_i)}$$

and let $\mathcal{D} = \{x_1, ..., x_n\}$ be a set of n samples independently drawn according to this probability density.

Question 8, Part (a)

Let $\mathsf{s} = (s_1, ..., s_d)^T$ be the sum of the n samples, show that

$$P(\mathcal{D}|\theta) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{(n-s_i)}$$

Question 8, Part (b)

Assuming a uniform a priori distribution over θ and using the identity

$$\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!}$$

show that

$$p(\theta|\mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$$

Question 8, Part (c)

Integrate the product $P(\mathrm{x}|\theta)p(\theta|\mathcal{D})$ over θ to obtain the desired conditional probability

$$P(\mathrm{x}|\mathcal{D}) = \prod_{i=1}^{d} \left(\frac{s_i + 1}{n + 2}\right)^{x_i} \left(1 - \frac{s_i + 1}{n + 2}\right)^{1 - x_i}$$

Question 8, Part (d)

If we think of obtaining $P(\mathsf{x}|\mathcal{D})$ by substituting an estimate $\hat{\theta}$ for θ in $P(\mathsf{x}|\theta)$, what is the effective Bayesian estimate for θ?

Question 9

What is the difference between Bayesian Estimation and MLE?