AVD613 - Machine Learning for Signal Processing

Assignment - 2 (Theory)

MLE and MAP

Submitted by

Vatsalya Gupta
SC19B098
B.Tech. ECE VII Semester

Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram - 695 547
October 2022

Problem 1: Maximum Likelihood Estimation

1. The likelihood of the uniform distribution is given by

$$L(a) = \prod_{i=1}^{n} p(x_i; a) = \begin{cases} 0 & \text{if } a < \max_{i=1}^{n} x_i \\ a^{-n} & \text{if } a \geq \max_{i=1}^{n} x_i \end{cases}$$

We know that the function a^{-n} is positive and decreasing with a, the parameter which will maximise the likelihood is the smallest a that is not less than $\max_{i=1}^{n} x_i$. Therefore

$$\hat{a}_{ML} = \max_{i=1}^{n} x_i \tag{1}$$

The log-likelihood of the exponential distribution is

$$l(\eta) = -n \log \eta - \frac{1}{\eta} \sum_{i=1}^{n} x_i$$

For maximising it, we calculate its derivative

$$\frac{dl}{d\eta} = -\frac{n}{\eta} + \frac{1}{\eta^2} \sum_{i=1}^{n} x_i$$

The derivative vanishes at $\eta = \frac{1}{n} \sum_{i=1}^{n} x_i$, and is positive for small values of η's, and negative for large values. Therefore

$$\hat{\eta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2}$$

is the parameter that maximizes the likelihood.

The log-likelihood of the normal distribution for a dataset is

$$l(\mu, \theta) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Since the log-likelihood approaches $-\infty$ if $\mu \to \pm\infty, \sigma \to 0$, or $\sigma \to \infty$, the log-likelihood is maximized for some finite $(\hat{\mu}_{ML}, \hat{\sigma}_{ML}), \hat{\sigma}_{ML} > 0$. At the maximum

$$\left(\frac{\partial l}{\partial \mu}(\hat{\mu}_{ML}, \hat{\sigma}_{ML}), \frac{\partial l}{\partial \sigma}(\hat{\mu}_{ML}, \hat{\sigma}_{ML}) \right) = 0$$

The first partial derivative $\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$ vanishes at $\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$. The second partial derivative $\frac{\partial l}{\partial \mu} = -\frac{n}{\sigma} + \sigma^{-3} \sum_{i=1}^{n} (x_i - \mu)^2$ vanishes when $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$. Thus

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad , \quad \hat{\sigma}_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \tag{3}$$

2. The expected value of $\hat{a}_{ML}$ is given by

$$\mathbb{E}_{p(x;a)}[\max\{X_1, X_2, \ldots, X_n\}] = \int_{0\leq x_1,x_2,\ldots,x_n\leq a} a^{-n}\left(\max_{i=1}^{n} x_i\right)\, dx_1\, dx_2\, \ldots\, dx_n \tag{4}$$

$\max_{i=1}^n x_i$ is always $\leq a$, and there are regions of positive probability on which $\max_{i=1}^n x_i$ is strictly less than a (for example $0 \leq x_1, x_2, \ldots, x_n \leq a/2$ occurs with positive probability). So, the expected value of $\max_{i=1}^n x_i$ must be strictly less than a, and the estimator is biased.

As a property of the exponential distribution

$$\mathbb{E}_{p(x;\eta)}[X] = \eta$$

Therefore, we can say that

$$\mathbb{E}[\hat{\eta}_{ML}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \eta \tag{5}$$

So, the estimator is unbiased.

Similarly, for a Gaussian distribution we have

$$\mathbb{E}_{p(x;\mu,\sigma)}[X] = \mu$$

And, it follows that

$$\mathbb{E}[\hat{\mu}_{ML}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mu \tag{6}$$

So, the estimator is unbiased.

3. Compute the bias of ML estimator of the variance by expanding the square under the sum

$$\mathbb{E}[\hat{\sigma}^2_{ML}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right]$$

$$\implies \mathbb{E}[\hat{\sigma}^2_{ML}] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{1\leq i<j\leq n} X_i X_j\right)\right]$$

$$\implies \mathbb{E}[\hat{\sigma}^2_{ML}] = \frac{1}{n}\left[\left(1 - \frac{1}{n}\right)\sum_{i=1}^{n}\mathbb{E}[X_i^2] - \frac{2}{n}\sum_{1\leq i\leq j\leq n}\mathbb{E}[X_i X_j]\right]$$

Since n samples are drawn independently, for $i \neq j$ we have $\mathbb{E}[XiXj] = \mathbb{E}[Xi]\mathbb{E}[Xj] = \mathbb{E}[X]^2$. Similarly, $\mathbb{E}[X_i^2]$ is same for all i. Therefore

$$\implies \mathbb{E}[\hat{\sigma}^2_{ML}] = \frac{n-1}{n}(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = \frac{n-1}{n}\sigma^2 \tag{7}$$

Hence, $\hat{\sigma}^2_{ML}$ is biased. Whereas, $\mathbb{E}\left[\left(\frac{n}{n-1}\right)\hat{\sigma}^2_{ML}\right] = \sigma^2$ will be the expectation of an unbiased estimator.

4. Write the $MSE(\hat{\theta})$ expression in terms of bias and variance

$$MSE(\hat{\theta}) = \mathbb{E}_{X_i \sim P(X|\theta)} \left[(\hat{\theta} - \theta)^2 \right] = \mathbb{E} \left[\left(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta \right)^2 \right]$$

$$= \mathbb{E} \left[\left(\hat{\theta} - \mathbb{E}[\hat{\theta}] \right)^2 \right] + 2\mathbb{E} \left[\hat{\theta} - \mathbb{E}[\hat{\theta}] \right] \left(\mathbb{E}[\hat{\theta}] - \theta \right) + \left(\mathbb{E}[\hat{\theta}] - \theta \right)^2 \tag{8}$$

The first term in the right-hand side is $var(\hat{\theta})$, the second term is 0 because $\mathbb{E} \left[\hat{\theta} - \mathbb{E}[\hat{\theta}] \right] = 0$, and the third term is $bias(\hat{\theta})^2$. Hence, $MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta})$.

To prove the second statement, we compute

$$MSE(\hat{\sigma}^2_{n-1}) - MSE(\hat{\sigma}^2_{ML}) = \mathbb{E}[(\hat{\sigma}^2_{n-1} - \sigma^2)^2 - (\hat{\sigma}^2_{ML} - \sigma^2)^2]$$

$$= \mathbb{E}[(\hat{\sigma}^2_{n-1})^2] - \mathbb{E}[(\hat{\sigma}^2_{ML})^2] - 2\sigma^2(\mathbb{E}[\hat{\sigma}^2_{n-1}] - \mathbb{E}[\hat{\sigma}^2_{ML}]) = \left(\frac{n^2}{(n-1)^2} - 1 \right) \mathbb{E}[(\hat{\sigma}^2_{ML})^2] - \frac{2}{n}\sigma^4$$

So, to show that $MSE(\hat{\sigma}^2_{n-1}) > MSE(\hat{\sigma}^2_{ML})$, we need to prove that

$$\mathbb{E}[(\hat{\sigma}^2_{ML})^2] > \frac{2(n-1)^2}{n(2n-1)}\sigma^4 \tag{9}$$

For a Gaussian distribution we can find $\mathbb{E}[(\hat{\sigma}^2_{ML})^2]$ directly

$$\mathbb{E}[(\hat{\sigma}^2_{ML})^2] = \mathbb{E} \left[\frac{1}{n^2} \left(\sum_{i=1}^{n} (X_i - \bar{X})^2 \right) \right] = \frac{1}{n^2} \mathbb{E} \left[\left(\frac{n-1}{n} \sum_{i=1}^{n} X_i^2 - \frac{2}{n} \sum_{1 \leq i < j \leq n} X_i X_j \right) \right]$$

$$= \frac{1}{n^3} \{ (n-1)^2 \mathbb{E}[X^4] + [2(n-1) + (n-1)^3] E[X^2]^2 + (n-1)(n-2)(n-3)\mathbb{E}[X]^4$$
$$-2(n-1)(n-2)(n-3)\mathbb{E}[X^2]\mathbb{E}[X]^2 - 4(n-1)^2\mathbb{E}[X^3]\mathbb{E}[X] \}$$

Because X is normal of mean μ and variance σ^2

$$\mathbb{E}[X] = \mu \quad , \quad \mathbb{E}[X^2] = \mu^2 + \sigma^2 \quad , \quad \mathbb{E}[X^3] = \mu^3 + 3\mu\sigma^2 \quad , \quad \mathbb{E}[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

$$\implies \mathbb{E}[(\hat{\sigma}_{ML})^2] = \left(1 - \frac{1}{n^2} \right) \sigma^4 \tag{10}$$

We can verify that $1 - \frac{1}{n^2} > \frac{2(n-1)^2}{n(2n-1)} \, \forall \, n > 1$, therefore $MSE(\hat{\sigma}^2_{n-1}) > MSE(\hat{\sigma}^2_{ML})$.

Problem 2: Maximum A-Posteriori Estimation

1. To calculate $\hat{\theta}_{ML}$ we maximise the log-likelihood on $[0, 1]$

$$l(\theta) = n_H \log\theta + (n - n_H)\log(1 - \theta)$$

We compute the derivative of the log-likelihood

$$\frac{dl}{d\theta} = \frac{n_H}{\theta} - \frac{n - n_H}{1 - \theta}$$

Solving for its zeroes, we get

$$\frac{n_H}{\theta} = \frac{n - n_H}{1-\theta} \implies \theta = \frac{n_H}{n} \tag{11}$$

If $0 < n_H < n$ then $l(\theta) \to -\infty$ if $\theta \to 0$ or $\theta \to 1$. It follows that the maximum is achieved on $(0,1)$ and $\hat{\theta}_{ML} = n_H/n$. The special cases $n_H = 0$ and $n_H = n$ result in the same formula for the estimator.

Also, we can apply Jensen's inequality on the conclave logarithm as

$$\frac{n_H}{n}\log\frac{\theta}{n_H} + \frac{n-n_H}{n}\log\left(\frac{1-\theta}{n-n_H}\right) \le \log\left[\frac{n_H}{n}\frac{\theta}{n_H} + \frac{n-n_h}{n}\frac{1-\theta}{n-n_H}\right] = \log\frac{1}{n}$$

The equality holds when

$$\frac{n_H}{\theta} = \frac{n-n_H}{1-\theta} \tag{12}$$

2. The prior has positive probability for only $\theta = 0.4$ or $\theta = 0.5$. So, to maximise the posterior we need to compare only two quantities

$$\hat{\theta}_{MAP} = \arg\max_{\theta\in\{0.4,0.5\}} \theta_{n_H}(1-\theta)^{n-n_H}0.5$$

$$\text{Therefore}\quad \hat{\theta}_{MAP} = 0.5 \quad \text{if} \quad 0.4^{n_H}\,0.6^{n-n_H} < 0.5^{n_H}\,0.5^{n-n_H} \implies 0.4^{n_H/n}\,0.6^{1-n_H/n} < 0.5$$
$$\implies \log\frac{0.6}{0.5} < \frac{n_H}{n}\log\frac{0.6}{0.4} \implies \frac{n_H}{n} > \frac{\log 1.2}{\log 1.5}$$

Hence, the MAP estimate is given by

$$\hat{\theta} = \begin{cases} 0.4 & \text{if } \frac{n_H}{n} < \frac{\log 1.2}{\log 1.5} \\ 0.5 & \text{if } \frac{n_H}{n} > \frac{\log 1.2}{\log 1.5} \end{cases} \tag{13}$$

The boundary case $\frac{n_H}{n} = \frac{\log 1.2}{\log 1.5}$ will not occur since the right-hand side is irrational.

3. The MAP estimate maximizes the following posterior

$$\hat{\theta}_{MAP} = \arg\max_{\theta\in[0,1]} \theta^{n_H+\alpha-1}(1-\theta)^{n-n_H+\beta-1}$$

We can see that the posterior is identical to the likelihood of the coin after observing exactly $n_H + \alpha - 1$ heads out of a total of $n + \alpha + \beta - 2$ tosses. As discussed in part 1

$$\hat{\theta}_{MAP} = n_H + \alpha - 1 n + \alpha + \beta - 2$$

$$\implies \lim_{n\to\infty}\frac{\hat{\theta}_{MAP}}{\hat{\theta}_{ML}} = \lim_{n\to\infty}\frac{1+(\alpha-1)/n_H}{1+(\alpha+\beta-2)/n} = 1 + \frac{\alpha-1}{\lim_{n\to\infty} n_H} = 1 \tag{14}$$

If $\theta = 0$ the last equality will not hold. In that case n_H must always be 0, so $\hat{\theta}_{MAP} = (\alpha-1)/(n+\alpha+\beta-2) \to 0 = \hat{\theta}_{ML}$.

4. $\hat{\theta}^1_{MAP}$ will be able to get closer to $\theta = 0.41$ with less number of samples since it only needs to rule out $\theta = 0.5$ out of the two feasible choices for the parameter θ. However, if the data is sufficiently more, $\hat{\theta}_{ML}$ is better because the MAP estimate will never be able to get closer than 0.01 to the true value of the parameter. Whereas, in the limit $\hat{\theta}_{ML}$ recovers 0.41 exactly.

Problem 3: MLE and MAP on Gaussian

1. The likelihood is given by

$$P(x_1,\ldots,x_N|\mu) = \prod_{i=1}^{N} P(x_i|\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)}$$

Since log is monotonically increasing, we can maximize the log-likelihood

$$\log(P(x_1,\ldots,x_N|\mu)) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2} \tag{15}$$

Equating the first derivative equal to zero, we get

$$\frac{d\log(P(x_1,\ldots,x_N|\mu))}{d\mu} = \sum_{i=1}^{N} \frac{(x_i-\mu)}{\sigma^2} = 0 \implies \sum_{i=1}^{N}(x_i-\mu) = 0$$

$$\implies \sum_{i=1}^{N} \mu = \sum_{i=1}^{N} x_i \implies N\mu = \sum_{i=1}^{N} x_i \implies \hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{16}$$

2. Using Bayes' rule to write the posterior probability

$$P(\mu|x_1,\ldots,x_n) = \frac{P(x_1,\ldots,x_N|\mu)P(\mu)}{P(x_1,\ldots,x_N)}$$

We can write that

$$P(x_1,\ldots,x_N|\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)} \quad , \quad P(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} e^{-(\mu-v)^2/(2\beta^2)}$$

So, using log-likelihood on Bayes' expression, we get

$$\log(P(\mu|x_1,\ldots,x_n)) = \left(\sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) - \log\left(\sqrt{2\pi\beta^2}\right) - \frac{(\mu-v)^2}{2\beta^2} \tag{17}$$

For maximizing log-likelihood, take derivative on both sides and equate to zero

$$\frac{\partial \log(P(\mu|x_1,\ldots,x_n))}{\partial\mu} = \left(\sum_{i=1}^{N} \frac{x_i-\mu}{\sigma^2}\right) - \frac{\mu-v}{\beta^2} = 0 \implies \frac{\mu-v}{\beta^2} = \sum_{i=1}^{N} \frac{x_i-\mu}{\sigma^2}$$

$$\implies \frac{\mu-v}{\beta^2} = \frac{\sum_{i=1}^{N} x_i}{\sigma^2} - \frac{N\mu}{\sigma^2} \implies \frac{\mu}{\beta^2} + \frac{N\mu}{\sigma^2} = \frac{\sum_{i=1}^{N} x_i}{\sigma^2} + \frac{v}{\beta^2}$$

$$\implies \frac{(\sigma^2+N\beta^2)\mu}{\sigma^2\beta^2} = \frac{\sigma^2 v + \beta^2\sum_{i=1}^{N} x_i}{\sigma^2\beta^2} \implies \hat{\mu} = \frac{\sigma^2 v + \beta^2\sum_{i=1}^{N} x_i}{\sigma^2+N\beta^2} \tag{18}$$

3. We know that the MLE and MAP estimators are given by

$$\hat{\mu}_{MLE} = \frac{\sum_{i=1}^{N} x_i}{N} \quad , \quad \hat{\mu}_{MAP} = \frac{\sigma^2 v + \beta^2\sum_{i=1}^{N} x_i}{\sigma^2+N\beta^2} = \frac{\sigma^2 v}{\sigma^2+N\beta^2} + \frac{\left(\sum_{i=1}^{N} x_i\right)/N}{1+\sigma^2/(N\beta^2)}$$

$$\implies \lim_{N\to\infty} \hat{\mu}_{MAP} = 0 + \frac{\left(\sum_{i=1}^{N} x_i\right)/N}{1+0} = \frac{\sum_{i=1}^{N} x_i}{N} = \hat{\mu}_{MLE} \tag{19}$$

Thus, as the number of samples increases, the two estimators become identical.

Problem 4: Parameter Estimation

1. Given that we have n i.i.d. samples. To find λ using MLE

$$P(D|\lambda) = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} \implies \hat{\lambda} = \arg\max_{\lambda} P(D|\lambda) = \arg\max_{\lambda} \log(P(D|\lambda))$$

For maximizing the log-likelihood, we can set the first derivative as zero

$$\log(P(D|\lambda)) = -n\lambda + \sum_{i=1}^{n} (X_i \log \lambda - \log(X_i!)) \implies \frac{d \log(P(D|\lambda))}{d\lambda} = -n + \sum_{i=1}^{n} \frac{X_i}{\lambda} = 0$$

$$\implies -n\lambda = \sum_{i=1}^{n} X_i \implies \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{20}$$

We also need to show the unbiased nature of this estimate

$$\mathbb{E}_{\lambda}[\hat{\lambda}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} \lambda = \lambda \tag{21}$$

In this expression, when the true parameter is λ, each X_i drawn from the Poisson distribution with parameter λ has $\mathbb{E}[X_i] = \lambda$. Since $\mathbb{E}_{\lambda}[\hat{\lambda}] - \lambda = 0$, we can say that the MLE of λ is unbiased.

2. We can assume the prior to be Gamma distribution, then the posterior distribution will be given by

$$P(\lambda|D) \;\propto\; P(D|\lambda) \cdot P(\lambda) = \left(\prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right) \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$= \lambda^{(\sum_{i=1}^{n} X_i + \alpha - 1)} e^{-(\lambda n + \beta\lambda)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \prod_{i=1}^{n} \frac{1}{X_i!} \;\propto\; \lambda^{\left(\sum_{i=1}^{n} X_i + \alpha - 1\right)} e^{-(\lambda n + \beta\lambda)}$$

$$\implies P(\lambda|D) \;\propto\; p\left(\lambda; \sum_{i} X_i + \alpha, n + \beta\right) \tag{22}$$

Here, $p\left(\lambda; \sum_i X_i + \alpha, n+\beta\right) = \Gamma\left(\sum_i X_i + \alpha, n+\beta\right)$. We can get the normalizing constant by using the law of total probability since

$$\int_{\lambda} c \cdot \lambda^{\left(\sum_{i=1}^{n} X_i + \alpha - 1\right)} e^{-(\lambda n + \beta\lambda)} = 1$$

Hence, the posterior distribution is also Gamma with the parameters $\Gamma\left(\lambda; \sum_i X_i + \alpha, n+\beta\right)$.

3. By definition, a probability distribution attains the point with the highest probability exactly at its mode. We know that the mode of $\Gamma(\alpha, \beta)$ is $(\alpha - 1)/\beta$ for $\alpha > 1$. If α is initially greater than 1, it is easy to see that $\alpha + \sum_i X_i > 1$, so the MAP estimate, given by the mode of $\Gamma(\sum_i X_i \alpha, n + \beta)$, is

$$\hat{\lambda}_{MAP} = \frac{\sum_i X_i + \alpha - 1}{n + \beta} \tag{23}$$

4. We write the log-likelihood as the following

$$L = \sum_i \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)}\right) = \sum_i \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

We need to maximize L, which we can do by setting $\partial L/\partial\mu$ to 0. This will give us the option **(c)** as the correct answer.

$$\mu_{MLE} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{24}$$

5. For MAP estimate, we maximize $f(\mu)f(X|\mu)$

$$f(\mu)f(X|\mu) = \frac{1}{\sigma_p\sqrt{2\pi}} e^{-(\mu-\mu_p)^2/(2\sigma_p^2)} \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)}$$

We maximize this with respect to μ after taking a logarithm, which will give

$$\frac{\sum_i x_i}{\sigma} + \frac{\mu_p}{\sigma_p} - \mu\left(\frac{N}{\sigma} + \frac{1}{\sigma_p}\right) = 0$$

Hence, solution will be option **(c)**, i.e.

$$\mu_{MAP} = \frac{\sigma^2 + \sigma_p^2 \sum_{i=1}^{N} x_i}{\sigma^2 + N\sigma_p^2} \tag{25}$$

6. The correct statements are **(b)**, **(c)** and **(d)**.

MAP estimates are equivalent to the ML estimates when the prior used in the MAP is a uniform prior over the parameter space.

One drawback of maximum likelihood estimation is that in some scenarios (hint: multinomial distribution), it may return probability estimates of zero.

The parameters which minimize the expected Bayesian L1 Loss is the median of the posterior distribution.

References

Various online resources and university courses were referred for gaining the necessary background knowledge to solve the assignment. Some of them are listed here.

1. Machine Learning (6.867), Massachusetts Institute of Technology, USA
2. Machine Learning (10-601), Carnegie Mellon University, USA
3. Optimization (10-725), Carnegie Mellon University, USA
4. Pattern Recognition and Machine Learning (CS5691), IIT Madras, India

■ ■ ■