Tutorial 3, Feb 2, 2024

Bayesian Estimation

Maximum a Posteriori (MAP) Estimation

- Consider random variables $\Theta, X_1, \ldots, X_n$ where $X_1, \ldots, X_n$ are conditionally independent given Θ
  - Our likelihood function is then $f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$
- Let the prior distribution for Θ be $f(\theta)$
- $\hat{\theta}_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}}\, f(\theta|x_1, \ldots, x_n) = \underset{\theta}{\operatorname{argmax}}\, \dfrac{f(\theta)f(x_1, \ldots, x_n|\theta)}{f(x_1, \ldots, x_n)} = \underset{\theta}{\operatorname{argmax}}\, f(\theta) \prod_{i=1}^{n} f(x_i|\theta)$
  - Just like MLE we can take the log to turn this into a sum
- Example: Let $W_1, \ldots, W_n$ be independent RVs each $W_i \sim \mathcal{N}(0, \sigma_i^2)$ and let $X_i = \Theta + W_i$; find the MAP and LMS estimates of Θ given a prior $\Theta \sim \mathcal{N}(\mu, \sigma^2)$
  - Given $\Theta = \theta$, $X_i \sim \mathcal{N}(\theta, \sigma_i^2)$
  - $$\begin{aligned} \hat{\theta}_{\text{MAP}} &= \underset{\theta}{\operatorname{argmax}} \left[\log f(\theta) + \sum_{i=1}^{n} \log(f(x_i|\theta)) \right] \\ &= \underset{\theta}{\operatorname{argmin}} \left[\frac{(\theta-\mu)^2}{2\sigma} + \sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\sigma_i^2} \right] \\ &= \frac{\frac{\mu}{\sigma^2} + \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\frac{1}{\sigma^2} + \sum_{i=1}^{n} \frac{1}{\sigma^2}} \end{aligned}$$
  - Note if we did MLE instead we would not have the σ^2 and μ terms in the expression
    - Using the prior for MAP is like having an additional measurement
  - As $n \to \infty$ this converges to the MLE estimate

Least Mean Square (LMS) Estimation

- Let the estimate for Θ be $g(X_1, \ldots, X_n)$, then we seek to minimize the expectation of the squared error of this estimate from the true parameter
  - $g^* = \underset{g}{\operatorname{argmin}}\, E[(g(X_1, \ldots, X_n) - \Theta)^2]$
- We can show that this is equal to the conditional expectation $g^*(\boldsymbol{X}) = E[\Theta|X_1 = x_1, \ldots, X_n = x_n]$
- $\hat{\theta}_{\text{LMS}} = E[\Theta|X_1 = x_1, \ldots, X_n = x_n] = \int \theta f(\theta|x_1, \ldots, x_n)\, \mathrm{d}\theta$
  - $f(\theta|x_1, \ldots, x_n)$ can be found using Bayes' rule
- The LMS estimator is always unbiased
- Note that for a Gaussian prior with Gaussian likelihood, the posterior $f(\theta|x_1, \ldots, x_n)$ will also be Gaussian, which means its max and its mean are at the same location
  - This means that the LMS (which is the expectation/mean) and the MAP (which is the max/mode) estimators are the same