MAP Estimation

lundi 18 avril 2022 15:42

Baye's Inteference

- Unknown – Estimate RV Θ
  - ↳ given distribution $p(\Theta)$ or $f(\Theta)$
- Observation – Data RV X
  - ↳ Observation model $p(x|\theta)$ or $f(x|\theta)$
- Given X, find best estimate $\hat{\Theta}$ of Θ

Additive Gaussian Noise Channel

$$\begin{tikzcd} & Z\sim N(0,\sigma^2) \arrow[d] & \\ \Theta \arrow[r] & \oplus \arrow[r] & Y \end{tikzcd}$$

Observation Model: $f(y|\theta)=f_Z(y-\theta)$

$$\Theta = \begin{cases} 1, & \text{prob. } p \\ -1, & \text{prob. } 1-p \end{cases} \rightarrow \text{Signal Transmitted}$$

$Y = \underbrace{\Theta + Z}_{\text{Indep.}}$ → Received Signal / observation

- ↳ Known: $p_\Theta(1)=p$, $p_\Theta(-1)=1-p$
- ↳ Posterior: $p(\Theta|y)$

Posterior

- Conditional prob.
- For fixed x, prob dist. in Θ
  - ↳ $\Sigma_\Theta p(\theta|x)=1$, $\int f(\theta|x)d\theta=1$
  - ↳ $p(\theta|x)=g(\theta)$
- For diff. $x_1, x_2, \ldots$, $p(\theta|x_1), p(\theta|x_2)$ are diff. dist. in Θ
  - ↳ $p(\theta|x_1)=g_1(\theta)$, $p(\theta|x_2)=g_2(\theta)$

MAP (Maximum a Posteriori)

- Given data $X=x$, best estimate of Θ = ?
- Best → Minimize prob. error $\mathbb{P}(\Theta\neq\hat{\Theta})$
- Optimal Estimator: MAP Estimator
  - ↳ $\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$
  - ↳ fix x, treat $p(\theta|x)$ as $g(\theta)$
  - ↳ find θ^* maximizing $g(\theta)$
    - ↳ $\theta^* \rightarrow \hat{\theta}_{MAP}(x)$
  - ↳ Repeat for all x

Illustration of the MAP estimator.

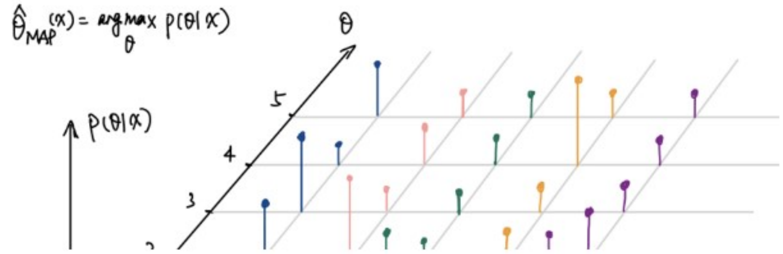
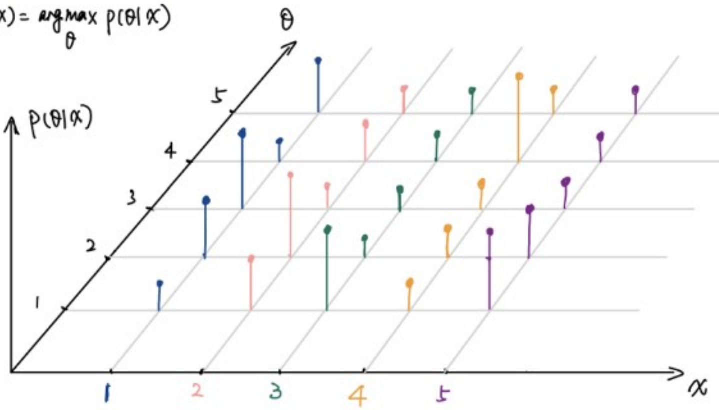


Illustration of the MAP estimator.

$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x)$



$\hat{\theta}_{MAP}(1)=3$ $\hat{\theta}_{MAP}(2)=2$ $\hat{\theta}_{MAP}(3)=1$ $\hat{\theta}_{MAP}(4)=4$ $\hat{\theta}_{MAP}(5)=1$.

$\hat{\theta}_{MAP}(x)$ is a function in x !

Maximum Likelihood (ML) estimation

- When distb. is **uniform** ($p(\theta) = c$ for all θ):

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} p(\theta|x) = \arg\max_{\theta} \frac{p(\theta)\,p(x|\theta)}{p(x)} = \arg\max_{\theta} \underbrace{p(x|\theta)}_{\text{Likelihood}} = \hat{\theta}_{ML}(x)$$

- ML Estimate : $\hat{\theta}_{ML}(x) = \arg\max_{\theta} p(x|\theta)$

Example: Infer the unknown bias in choosing lab problems.

- Θ is drawn uniform at random in the interval $[0,1]$.
- Given $\Theta=\theta$, each problem is chosen to be graded with probability θ. independent of each other
- There are n problems in total (fixed). K were chosen to be graded.
- Find the MAP estimator $\hat{\theta}_{MAP}(k)$

- Unknown $\Theta \sim Unif[0,1]$.
- Observation model: for $i=1,2,\cdots,n$, denote $X_i = \begin{cases} 1, & \text{if problem } i \text{ is graded.} \\ 0, & \text{o.w.} \end{cases}$

  $X_i|\Theta=\theta \sim Bern(\theta)$. $K = X_1+X_2+\cdots+X_n$. $K|\Theta=\theta \sim Binom(n,\theta)$

- Θ is uniform $\Rightarrow$ $\hat{\theta}_{MAP}(k) = \hat{\theta}_{ML}(k) = \arg\max_{\theta} p(k|\theta) = \arg\max_{\theta} \binom{n}{k} \underbrace{\boxed{\theta^k(1-\theta)^{n-k}}}_{\triangleq g(\theta)}$ $\longrightarrow$ $\arg\max_{\theta} \dfrac{p(k|\theta)f(\theta)}{p(k)} = \arg\max_{0\le\theta\le1} \dfrac{\binom{n}{k}\theta^k(1-\theta)^{n-k}}{p(k)}$

- Find maximum: $g'(\theta) = k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1} \overset{set}{=} 0 \Rightarrow \hat{\theta}_{MAP}(k) = \frac{k}{n}$.

  ↳ take derivative

Example: Estimating Gaussian signal in Gaussian noise.

- Signal $\Theta \sim N(0,1)$
- Observation $X = \Theta + W$, $W \sim N(0,1)$ indep. of Θ. Find $\hat{\theta}_{MAP}(x)$.

- Unknown: $\Theta \sim N(0,1)$ conditional pdf

  recall our trick in BSC example.

- Observation model: $f(x|\theta) = \underbrace{f_W(x-\theta)}_{f_{W|\Theta}(x-\theta|\theta)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$ → density of std normal

- Posterior: $f(\theta|x) = \dfrac{f(\theta)f(x|\theta)}{f(x)} = \dfrac{1}{f(x)} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} \cdot \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \triangleq c(x)\, e^{-\frac{1}{2}(\theta^2+(x-\theta)^2)}$

$\Rightarrow$ Want $\min_{\theta} [\theta^2 + (x-\theta)^2]$ for fixed x.

$$\frac{\partial[\theta^2+(x-\theta)^2]}{\partial\theta} \triangleq 0 \Rightarrow \theta^* = \frac{x}{2}. \Rightarrow \hat{\theta}_{MAP}(x) = \frac{x}{2}.$$

→ $\arg\max_{\theta} f(\theta|x) = \arg\max_{\theta}$

$= \arg\max_{\theta}$

Computing $\mathbb{P}(\hat{\Theta} \neq \Theta)$

$$\Theta \longrightarrow \boxed{P(x|\Theta)} \xrightarrow{x} \boxed{\hat{\Theta}(x)} \xrightarrow{\hat{\Theta}}$$

$\hat{\Theta} = \hat{\Theta}(x)$ → a number

$$\Theta \longrightarrow \boxed{P(x|\theta)} \xrightarrow{x} \boxed{\hat{\theta}(x)} \xrightarrow{\hat{\theta}}$$

$\hat{\Theta} = \hat{\Theta}(x) \rightarrow$ a number

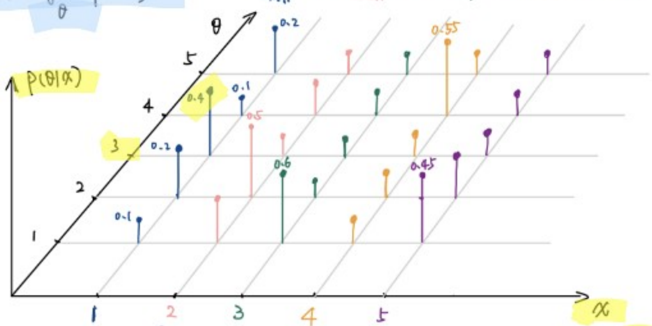
$$\mathbb{P}(\hat{\Theta} \neq \Theta) = \sum_x P_x(x)\, \mathbb{P}(\hat{\Theta} \neq \Theta \,|\, X = x) = \sum_x P_x(x)\, \mathbb{P}(\hat{\theta}(x) \neq \Theta \,|\, X = x)$$

- Conditional prob. error given data x: $\mathbb{P}(\hat{\Theta} \neq \Theta \,|\, X = x)$

Illustration of the probability of error.

$\hat{\theta}_{MAP}(x) = \underset{\theta}{\arg\max}\; p(\theta|x)$

$\hat{\theta}_{MAP}(1) = 3$ $\hat{\theta}_{MAP}(2) = 2$ $\hat{\theta}_{MAP}(3) = 1$ $\hat{\theta}_{MAP}(4) = 4$ $\hat{\theta}_{MAP}(5) = 1$.



$\mathbb{P}(\hat{\Theta} \neq \Theta | X=1) = P(\hat{\theta}_{MAP}(1) \neq \Theta | X=1) = \mathbb{P}(3 \neq \Theta | X=1) = 1 - \mathbb{P}(\Theta = 3 | X=1) = 1-0.4 = 0.6$

$\mathbb{P}(\hat{\Theta} \neq \Theta | X=2) = 1-0.5 = 0.5$ $\mathbb{P}(\hat{\Theta} \neq \Theta | X=3) = 1-0.6 = 0.4$ $\mathbb{P}(\hat{\Theta} \neq \Theta | X=4) = 1-0.55 = 0.45$ $\mathbb{P}(\hat{\Theta} \neq \Theta | X=5) = 1-0.45 = 0.55$

$\mathbb{P}(\Theta \neq \Theta) = 0.6\, P_X(1) + 0.5\, P_X(2) + 0.4\, P_X(3) + 0.45\, P_X(4) + 0.55\, P_X(5)$

- MAP Estimator Minimizes $\mathbb{P}(\hat{\Theta} \neq \Theta)$

Additive Gaussian Noise

$$\Theta \longrightarrow \oplus \longrightarrow Y, \quad Z \sim N(0, \sigma^2) \downarrow$$

- Binary RV $\Theta = \begin{cases} 1, & \text{prob } 1/2 \\ -1, & \text{prob } 1/2 \end{cases}$
- Observation $Y = \underline{\Theta + Z} \rightarrow$ indep.
- Find $\hat{\Theta}$ minimizing $\mathbb{P}(\hat{\Theta} \neq \Theta)$ & $\mathbb{P}(\hat{\Theta} \neq \Theta)$:
  - ↳ Use MAP
  - ↳ prior is uniform: $p_\Theta(-1) = p_\Theta(1) = 1/2$
  - $\therefore \hat{\Theta}_{MAP}(y) = \hat{\Theta}_{ML}(y)$

$$\hat{\Theta}_{ML}(y) = \underset{\theta}{\arg\max}\, f(y|\theta) = \underset{\theta}{\arg\max}\, f_Z(y-\theta) = \begin{cases} 1, & y > 0 \\ -1, & y \leq 0 \end{cases}$$

$\cdot\ \mathbb{P}(\hat{\Theta} \neq \Theta) = \sum_\theta \mathbb{P}(\hat{\Theta} \neq \theta, \Theta = \theta) = \mathbb{P}(\hat{\Theta} = -1, \Theta = 1) + \mathbb{P}(\hat{\Theta} = 1, \Theta = -1)$ → Probability of Error

$= \mathbb{P}(\Theta = 1, Y \leq 0) + \mathbb{P}(\Theta = -1, Y > 0) = \mathbb{P}(\Theta = 1)\mathbb{P}(Y \leq 0 | \Theta = 1) + \mathbb{P}(\Theta = -1)\mathbb{P}(Y > 0 | \Theta = -1)$

$= \frac{1}{2}\int_{-\infty}^{0} f_{Y|\Theta}(y|1)\, dy + \frac{1}{2}\int_{0}^{\infty} f_{Y|\Theta}(y|-1)\, dy$

$= \frac{1}{2}\phi(-\frac{1}{\sigma}) + \frac{1}{2}(1 - \phi(\frac{1}{\sigma}))$

$= 1 - \phi(\frac{1}{\sigma})$

$\phi(z) \triangleq \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$