MAXIMUM A POSTERIORI PRINCIPLE (MAP)

- BAYESIAN IN NATURE
- GOAL IS TO ESTIMATE PARAMETER θ OF A DISTRIBUTION $P(\theta)$ FROM A SET OF DATA POINTS $\mathcal{D}$.
- $\hat{\theta}_{MAP} = \underset{\theta}{\text{ARG MAX}}\ P(\theta | \mathcal{D})$
- $P(\theta | \mathcal{D})$ = POSTERIOR BELIEF, GIVEN OBSERVATIONS
  DATA
- $\hat{\theta}_{MAP} = \underset{\theta}{\text{ARG MAX}}\ P(\mathcal{D} | \theta)\, P(\theta)$

  WHERE $P(\theta)$ = PRIOR BELIEF

OBSERVATION $\quad \hat{\theta}_{MAP} = \underset{\theta}{\text{ARG}} \text{ MAX } P(\mathcal{D}|\theta)P(\theta)$

PROOF: $\quad \hat{\theta}_{MAP} = \underset{\theta}{\text{ARG}} \text{ MAX } P(\theta|\mathcal{D})$

$$= \underset{\theta}{\text{ARG}} \text{ MAX } \frac{P(\mathcal{D}|\theta)P(\theta)}{P(\mathcal{D})}$$

INDEPENDENT OF θ

$$= \underset{\theta}{\text{ARG}} \text{ MAX } P(\mathcal{D}|\theta)P(\theta) \quad \square$$

LET $\mathcal{D} = \{x_1, x_2, \ldots, x_m)$, THEN

$$\hat{\theta}_{MAP} = \underset{\theta}{\text{ARG}} \text{ MAX } P(x_1, x_2, \ldots, x_m|\theta)P(\theta)$$

IF OBSERVATIONS ARE INDEPENDENT, THEM

$$P(x_1, x_2, \ldots, x_m|\theta) = \prod_{i=1}^{m} P(x_i|\theta)$$

AS LOGARITHM IS A MONOTONICALLY INCREASING FUNCTION

$$\hat{\theta}_{MAP} = \underset{\theta}{\text{ARG}} \text{ MAX } \ln\left\{\prod_{i=1}^{m} P(x_i|\theta) \cdot P(\theta)\right\}$$

$$= \underset{\theta}{\text{ARG}} \text{ MAX } \Big(\underbrace{\ln P(\theta)}_{\text{EXTRA TERM}} + \underbrace{\sum_{i=1}^{m} \ln P(x_i|\theta)}_{\text{TERM OCCURS IN } \hat{\theta}_{ML}}\Big)$$

EXAMPLE

BERNOULLI DISTRIBUTION: $f_X(x; p) = p^x (1-p)^{1-x}$; $p > 0$; $x = 0, 1$

BETA DISTRIBUTION:

$$f_{BETA}(p; a, b) = \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1} \; ; \; 0 < p < 1 \; ; \; a, b > 0$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \text{BETA FUNCTION}$$

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \; ; \; Re(t) > 0 \; ; \; = \text{GAMMA FUNCTION}$$

$$\Gamma(n) = (n-1)! \; ; \; n \in \mathbb{P}$$

$\mathbb{P} = \{1, 2, 3, \dots\}$
= SET OF POSITIVE INTEGERS

$\Gamma(\cdot)$ FUNCTION IS A GENERALIZATION OF FACTORIAL FUNCTION FOR POSITIVE REAL NUMBERS

$\mathcal{D} = \{x_1, x_2, \dots, x_n\}$ = DATA SET

$$P(x_i | p) = p^{x_i} (1-p)^{1-x_i}$$

$$P(p) = \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1}$$

$$P(p | \mathcal{D}) \propto P(\mathcal{D} | p) P(p)$$

$\propto$ PROPORTIONAL

THAT IS, $P(p|\mathcal{D}) \propto \left\{ \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} \right\} p^{a-1} (1-p)^{b-1}$

$$\hat{p}_{MAP} = \underset{p}{\text{ARG MAX}} \underbrace{\left\{ \sum_{i=1}^{n} \left(x_i \ln p + (1-x_i) \ln(1-p) \right) + (a-1) \ln p + (b-1) \ln(1-p) \right\}}_{\mathcal{L}}$$

$$\frac{\partial \mathcal{L}}{\partial p} = \sum_{i=1}^{n} \left\{ \frac{x_i}{p} + \frac{(1-x_i)(-1)}{(1-p)} \right\} + \frac{(a-1)}{p} + \frac{(b-1)(-1)}{(1-p)} = 0$$

LET $S \triangleq \sum_{i=1}^{n} x_i$

$$\therefore \frac{S}{p} - \frac{(n-S)}{(1-p)} + \frac{(a-1)}{p} + \frac{(b-1)}{(1-p)}(-1) = 0$$

$$(S + a-1)(1-p) - p(n-S+b-1) = 0$$

$$(S+a-1) = p(n-S+b-1+S+a-1)$$

$$p = \frac{S+a-1}{n+a+b-2}$$

$$\therefore \hat{p}_{MAP} = \frac{\sum_{i=1}^{n} x_i + a-1}{n+a+b-2}$$

NOTE THAT, IF $n \to \infty$; THEN $\hat{p}_{MAP} \to \frac{1}{n}\sum_{i=1}^{n} x_i$

$$= \hat{p}_{MLE}$$

$\square$

EXAMPLE

NORMAL DISTRIBUTION: $$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\left(\frac{x-\mu}{\sigma}\right)^2\right];\ x \in \mathbb{R}$$

$$\mu \in \mathbb{R};\quad \sigma > 0$$

PRIOR DISTRIBUTION: THIS DISTRIBUTION IS ALSO NORMAL: WITH MEAN $= \nu$; AND VARIANCE β^2

$\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ = DATA SET

$$P(x_i \mid \mu, \sigma^2) = f_X(x_i;\mu,\sigma^2)$$

$$P(\mu) = \frac{1}{\sqrt{2\pi}\,\beta}\exp\left[-\frac{1}{2}\left(\frac{\mu-\nu}{\beta}\right)^2\right] \longleftarrow \text{PRIOR}$$

$$P(\mu \mid \mathcal{D}) \propto P(\mathcal{D} \mid \mu)\,P(\mu)$$

$$= \left\{\prod_{i=1}^{n} f(x_i;\mu,\sigma^2)\right\}\frac{1}{\sqrt{2\pi}\,\beta}\exp\left[-\frac{1}{2}\left(\frac{\overset{\mu}{\cancel{x}}-\nu}{\beta}\right)^2\right]$$

$$\hat{\mu}_{MAP} = \underset{\mu}{\text{ARG MAX}}\left\{\sum_{i=1}^{n}\ln f(x_i;\mu,\sigma^2) + \ln\frac{1}{\sqrt{2\pi}\,\beta} - \frac{1}{2}\left(\frac{\mu-\nu}{\beta}\right)^2\right\}$$

THIS IS EQUIVALENT TO MINIMIZING THE FOLLOWING FUNCTION OF μ.

$$g(\mu) = \left(\frac{\mu-\nu}{\beta}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$g'(\mu) = \frac{2}{\beta^2}(\mu-\nu) + \frac{2}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)(-1) = 0$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\beta^2\sum_{i=1}^{n}x_i + \sigma^2\nu}{n\beta^2+\sigma^2}$$

$$g''(\mu) = \frac{2}{\beta^2} + \frac{2}{\sigma^2}\cdot n > 0$$

NOTE THAT, IF $n \longrightarrow \infty$, THEN $\hat{\mu}_{MAP} \rightarrow \frac{1}{n}\sum_{i=1}^{n}x_i$

$$= \hat{\mu}_{MLE}$$

□