MLE, MAP, & NB

MLE = MAXIMUM LIKELIHOOD ESTIMATION

MAP = MAXIMUM A POSTERIORI

NB = NAIVE BAYES'

$\mathcal{D}$ = SET OF DATA GENERATED FROM SOME DISTRIBUTION PARAMETRIZED BY θ.

WE WANT TO ESTIMATE θ

MLE & MAP ARE PARAMETER ESTIMATION METHODS.

MAXIMUM LIKELIHOOD ESTIMATION (MLE)

IN THE MLE METHOD, WE WANT TO FIND θ THAT BEST EXPLAINS THE DATA.

$\Rightarrow$ WE MAXIMIZE $P(\mathcal{D} \mid \theta)$

DENOTE SUCH VALUE BY $\hat{\theta}_{ML}$

$$\hat{\theta}_{ML} = \arg\max_{\theta} P(\mathcal{D} \mid \theta)$$

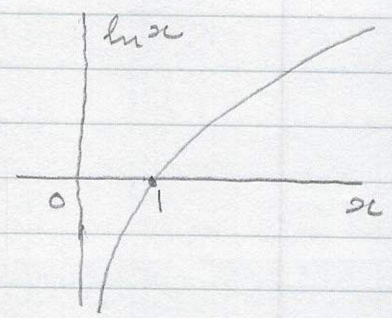
IF THE SET OF DATA POINTS $\mathcal{D} = (x_1, x_2, \ldots, x_n)$, THEN

$$\hat{\theta}_{ML} = \arg\max_{\theta} P(x_1, x_2, \ldots, x_n \mid \theta)$$

IF OBSERVATIONS ARE INDEPENDENT: $\cancel{P(x_1, x_2, \ldots, x_n) = \prod}$

$$P(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} P(x_i \mid \theta)$$

AS LOGARITHM IS A MONOTONICALLY INCREASING FUNCTION.



$$\hat{\theta}_{ML} = \arg\max_{\theta} \ln \prod_{i=1}^{n} P(x_i \mid \theta)$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \ln P(x_i \mid \theta)$$

IF WE KNOW $P(\cdot \mid \cdot)$ WE CAN SOLVE THIS BY TAKING DERIVATIVES WITH RESPECT TO θ AND MAKING IT EQUAL TO ZERO.

MAXIMUM A POSTERIORI (MAP)

θ = ASSUMPTIONS (MODE, HYPOTHESIS)

$\mathcal{D}$ = OBSERVATIONS (DATA)

$P(\theta|\mathcal{D})$ = POSTERIOR DISTRIBUTION
= POSTERIORI BELIEF GIVEN OBSERVATIONS

$P(\mathcal{D}|\theta)$ = LIKELIHOOD OF OBSERVATIONS, GIVEN THE MODEL

$P(\theta)$ = PRIOR DISTRIBUTION
= PRIOR BELIEF

$P(\mathcal{D})$ = MARGINAL LIKELIHOOD
= EVIDENCE : USED AS A NORMALIZATION FACTOR

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)\,P(\theta)}{P(\mathcal{D})} \; ; \; P(\mathcal{D}) > 0$$

LIKELIHOOD OF DATA, GIVEN MODEL

PRIOR BELIEF

$$P(\text{MODEL}|\text{DATA}) = \frac{P(\text{DATA}|\text{MODEL})\,P(\text{MODEL})}{P(\text{DATA})}$$

EVIDENCE

POSTERIOR BELIEF
GIVEN DATA

WE WANT TO FIND MOST LIKELY VALUE OF θ GIVEN $\mathcal{D}$.

THIS IS $\underset{\theta}{\text{ARG MAX}}\; P(\theta|\mathcal{D})$

MAXIMUM A-POSTERIORI (MAP) ESTIMATE IS DEFINED AS

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

MODEL DATA

NAIVE BAYES' CLASSIFIER

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ = SET OF LABELS FOR DIFFERENT CLASSES OF DATA

$P(\omega_i)$ = PROBABILITY THAT A DATA POINT BELONGS TO CLASS $\omega_i \in \Omega$

= A PRIORI PROBABILITY

$\mathcal{R}$ = FEATURE SPACE = SET OF ALL DATA POINTS

$\mathcal{X}$ = SET OF DATA POINTS, $\mathcal{X} \subseteq \mathcal{R}$

$\tilde{x}$ = RANDOM DATA VECTOR

$x \in \tilde{x}$ IS AN INSTANCE OF RANDOM DATA VECTOR

$p(x)$ = PROBABILITY DENSITY FUNCTION OF THE RANDOM DATA VECTOR $\tilde{x}$

$$p(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)} \; ; \; p(x) > 0 \; ; \; 1 \le i \le m$$

$$p(x) = \sum_{i=1}^{m} p(x|\omega_i)P(\omega_i)$$

TEST DATA POINT $x \in \mathcal{R} \setminus \mathcal{X}$ IS ASSIGNED TO CLASS ω_i

IF $P(\omega_i|x) > P(\omega_j|x) \quad \forall j \ne i \; ; \; 1 \le j \le m$

THIS SCHEME MINIMIZES CLASSIFICATION ERROR PROBABILITY.

THIS IS CALLED NB CLASSIFIER, BECAUSE IF $x_a \in \mathcal{R}$ IS $(x_{a_1}, x_{a_2}, \ldots, x_{a_t})$ IT IS ASSUMED THAT

$$p(x_a|\omega_j) = \prod_{k=1}^{t} p(x_{ak}|\omega_j) \; ; \quad 1 \le j \le m$$

DISCUSSION

1. MLE, MAP, AND NB ARE ALL CONNECTED
   - MLE & MAP ARE PARAMETER ESTIMATION METHODS.
   - NB IS A CLASSIFIER THAT PREDICTS THE PROBABILITY OF THE CLASS THAT THE EXAMPLE BELONGS TO.

2. MLE DOES NOT ALLOW US TO 'INJECT' OUR BELIEFS ABOUT THE LIKELY VALUES FOR THE PARAMETER (PRIOR) IN THE ESTIMATION PROCESS.

3. MAP ALLOWS THE FACT THAT THE PARAMETER CAN TAKE VALUES FROM A PRIORI (NON-UNIFORM) DISTRIBUTION THAT EXPRESS OUR PRIOR BELIEFS REGARDING THE PARAMETERS.

4. MAP RETURNS PARAMETER VALUE, WHERE THE PROBABILITY IS HIGHEST FOR GIVEN DATA.

5. MLE AND MAP, EACH RETURN A SINGLE SPECIFIC VALUE FOR THE PARAMETER.

   NB COMPUTES THE FULL POSTERIOR DISTRIBUTION $P(\theta|\mathcal{D})$.

OPINION

- MLE CAN BE REGARDED AS TRADITIONAL "FREQUENTIST" THINKING
- MAP CAN BE REGARDED AS BAYESIAN (AS IT IS A DIRECT APPLICATION OF BAYES' THEOREM).