4. Inference

Bayes - Posterior - MPE - MAP

Evidence

Sampling

5. Exact Inference

Bayes Filk

HMM \rightarrow FACTOR GRAPHS

\text{MAX PRODUCT F}_{n \text{ MPE}}

\text{SUM PRODUCT F}_{n \text{- posterior}} - \text{MAP}


diagram:

\text{Posterior} \quad P(X | A, O, X_0)

\rightarrow 625 \text{ numbers}

\frac{5^{100}}{5^{100}} = 8 \times 10^{49}

\text{Max Probable Explanation}

\rightarrow |X| = 5^4

= |X| \text{T}^T

= 625
Importance Sampling:

\[ \pi(x) = P(x|A, 0) \quad \text{TARGET POSTERIOR} \]

\[ q(x) = P(x|A) \quad \text{PROPOSAL} \]

\[ w^{(i)} = \frac{\pi(x^{(i)})}{q(x^{(i)})} \quad \text{importance weight} \]

over sample \(< 1.0 \]

under sample \(> 1.0 \]

\[ w^{(i)} = \frac{P(x|A, 0)}{P(x|A)} \]

Do N times:

1) Sample \( x^{(i)} \) from Markov Ch. \( \leq \)

2) "Score" with \( w^{(i)} = P(0|x) \)

\[ \{ \]

\[ \quad = P(0|x) = \prod P(0_{i+1}|x_t) \]

i.e. weight equals probability of measurements given actions taken = very intuitive?

Downside: many weights \( w^{(i)} \) will be close to 0: variance

Full Joint vs Weighted Sample

\[ \text{625} \quad \text{vs} \quad \text{1M 100} \]
\[ F(x) = \frac{x}{5} \]

\[ f(x) = \frac{\#E}{100} \]

\[ = \frac{\sum_{i=1}^{100} \#E(x)}{100} \]

Monte Carlo estimate of any \( f(x) \)

High variance if \( |x| \gg \)

**Bayes' Law**

\[ P(x, y) = P(x|y)P(y) = P(y|x)P(x) \]

prior

\[ P(x|y) = \frac{P(y|x)P(x)}{P(y)} \]

post.

\[ P(x|y) \propto P(y|x)P(x) \] likelihood

prior of \( x \) given \( y \)

\[ P(x|y) \propto L(x; y)P(x) \] posterior

\[ L(x; y) \propto P(y|x) \] likelihood

prior
Exact Inference

Discrete inference over Sequences

A = K(x) [60]

Hidden state sequence

HIDDEN STATE SEQUENCE

A

Hidden Markov model

P(x_t | O^t) = P(x_t | O_t, O^{t-1})

α L(x_0 | O_t) P(x_0 | O^{t-1})

α L(x_t; O_t) \sum P(x_t | x_{t-1}) P(x_{t-1} | O^{t-1})

Filtering distribution

α L(x_t; O_t) P(x_t | x_{t-1}) P(x_{t-1} | O^{t-1})

1) calculate predictive: → predict
2) multiply with likelihood: → update