Izmir Institute of Technology

CENG 461 – Artificial Intelligence

Hidden Markov Models and Filters

2015-16 Fall Term

Temporal Sequence of Random Variables

- Hidden Markov Models (HMMs) are used to analyze or to predict time series, i.e. temporal sequences of random variables.
- We can assume that a sequence of random variables are all independent from each other.
- This yields a very efficient model in terms of computation, but it is weak to represent relations.

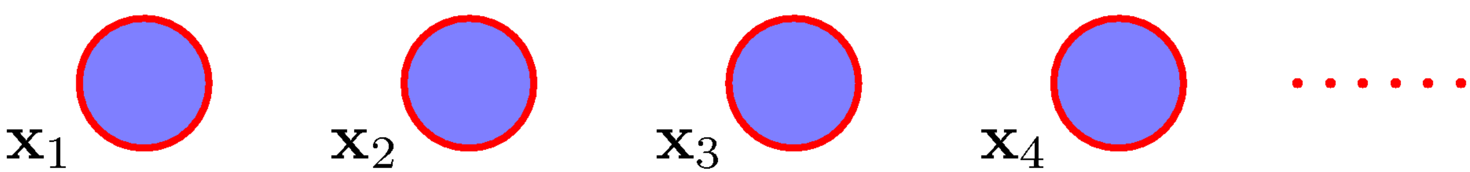


Figure from Pattern Recognition and Machine Learning by C.M. Bishop

Temporal Sequence of Random Variables

- A better model is to condition each random variable only on the previous one:

$$P(x_n|x_1, x_2, \dots, x_N) = P(x_n|x_{n-1})$$

- Such a sequence is called a Markov chain.

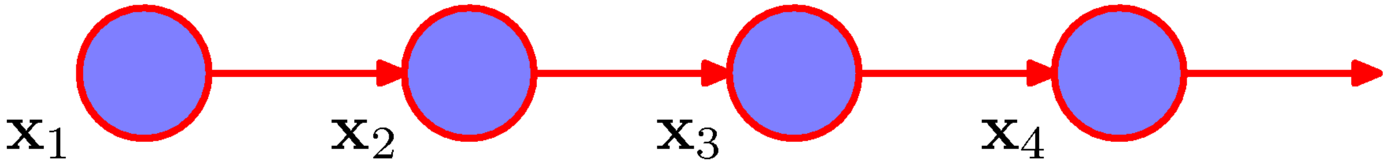


Figure from Pattern Recognition and Machine Learning by C.M. Bishop

Temporal Sequence of Random Variables

- We can increase the complexity by modelling dependencies between more and more elements:

$$P(x_n|x_1, x_2, \dots, x_N) = P(x_n|x_{n-1}, x_{n-2})$$

- Example: A second order Markov chain

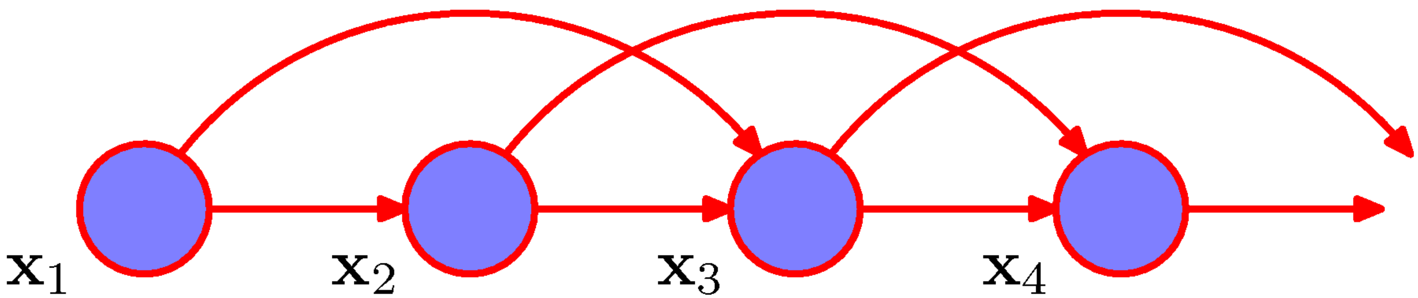
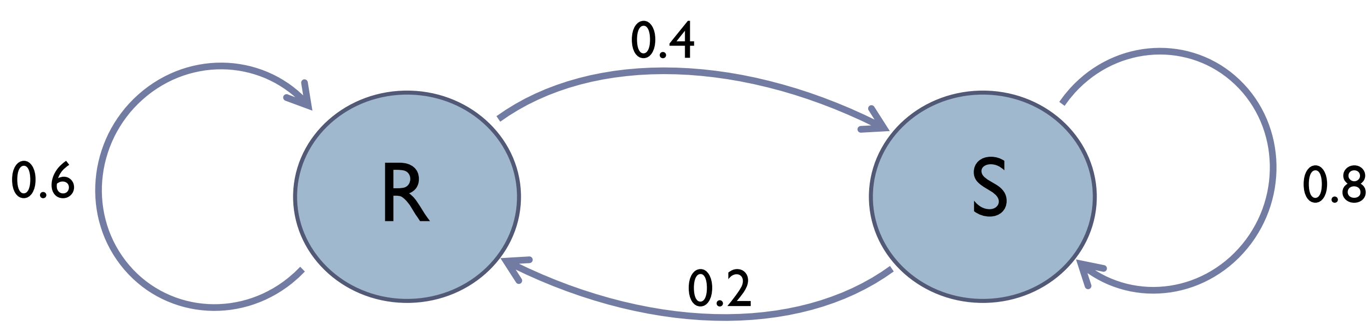


Figure from Pattern Recognition and Machine Learning by C.M. Bishop

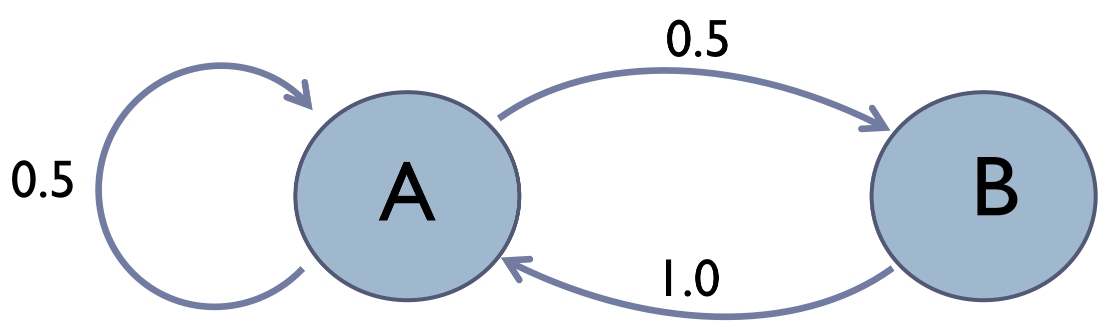
Markov Chain Example

- Let us represent a Markov chain using a state transition diagram. Each time step can take two values: R or S.
- R represents Rainy. S represents Sunny. Transition probabilities (from time t to t+1) are written on the edges.
- Starting probabilities: $P(R_0)=1$ and $P(S_0)=0$
- $P(R_1)=?$ 0.6
- $P(R_2)=?$ 0.44
- $P(R_3)=?$ 0.376



Markov Chain - Stationary Distribution

- Stationary distribution is the probabilities after a huge number of time steps, actually when $t \rightarrow \infty$.
- For the below state transition diagram, $P(A_\infty)= ?$
- It happens when $P(A_t)=P(A_{t-1})$.
- I.e. $P(A_t|A_{t-1})\ P(A_{t-1})+ P(A_t|B_{t-1})\ P(B_{t-1}) = P(A_{t-1})$.
- $P(A_\infty)= 0.666$



Introducing Hidden Variables

- Sometimes we can not observe the actual states (z_n), instead we observe some variables (x_n), related to the hidden variables.

$$P(x_1, x_2, \ldots, x_N, z_1, z_2, \ldots, z_N)$$
$$= P(z_1) \prod_{n=2}^{N} P(z_n | z_{n-1}) \prod_{n=1}^{N} P(x_n | z_n)$$

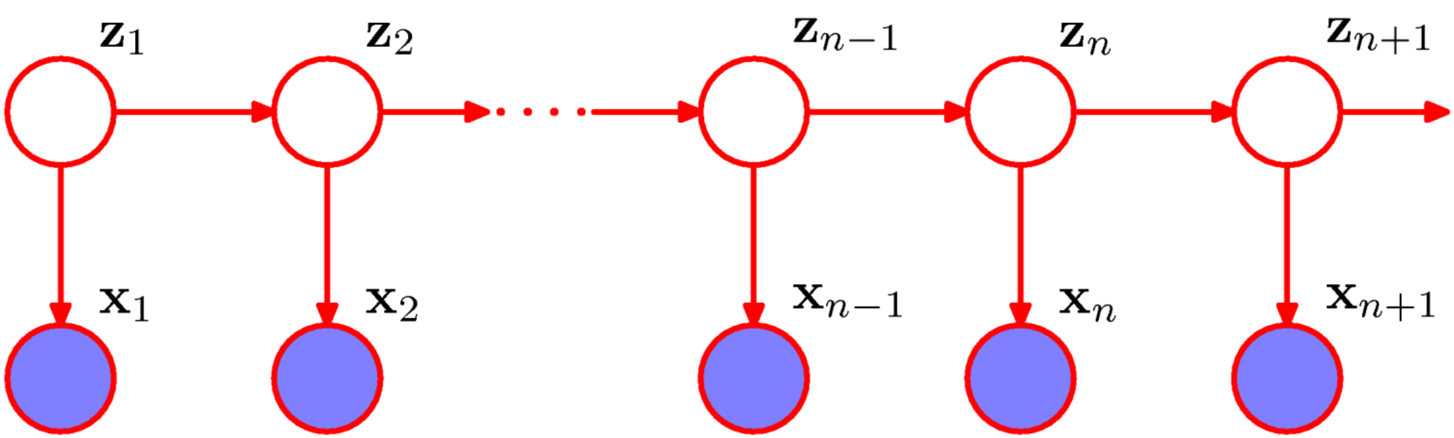


Figure from Pattern Recognition and Machine Learning by C.M. Bishop

Introducing Hidden Variables

- If the hidden variables are discrete, this type of model is called a Hidden Markov Model, the observables might be discrete or continuous.
- Hidden variables are also called as latent variables.

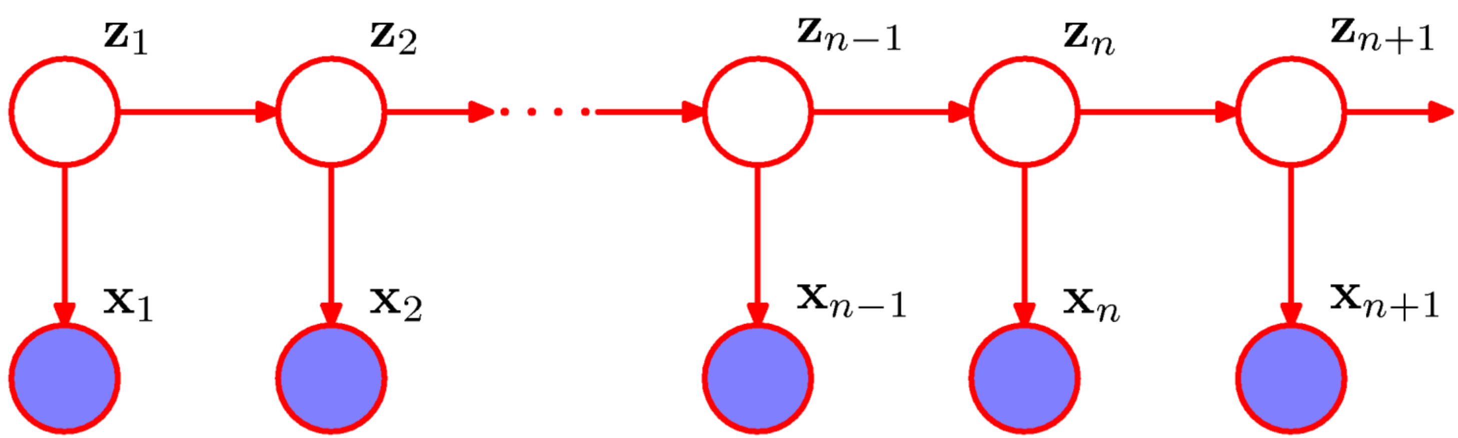
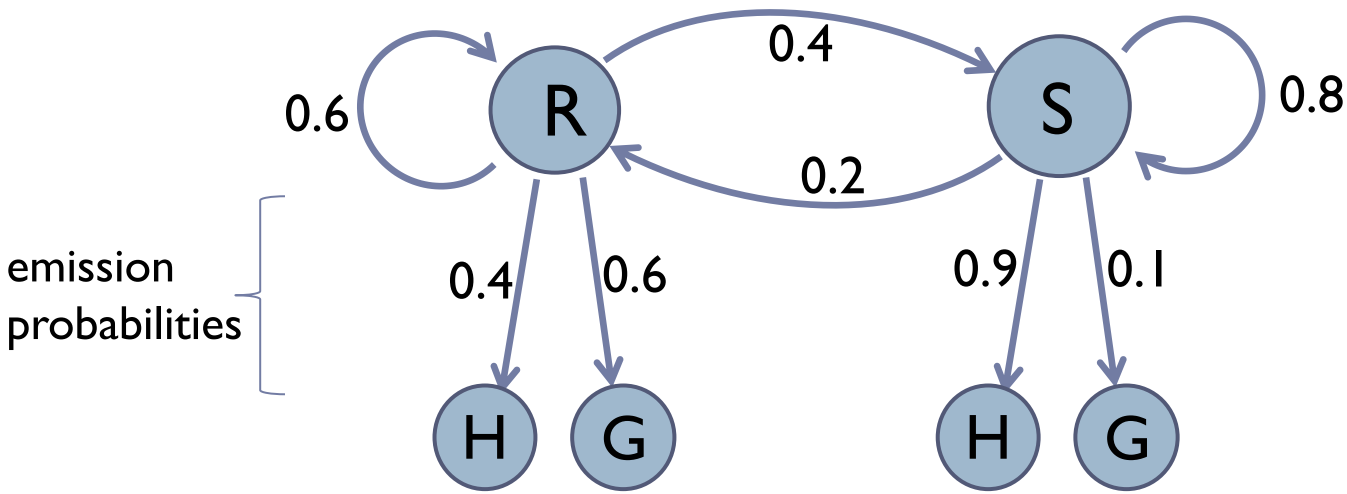


Figure from Pattern Recognition and Machine Learning by C.M. Bishop

Hidden Markov Model Example

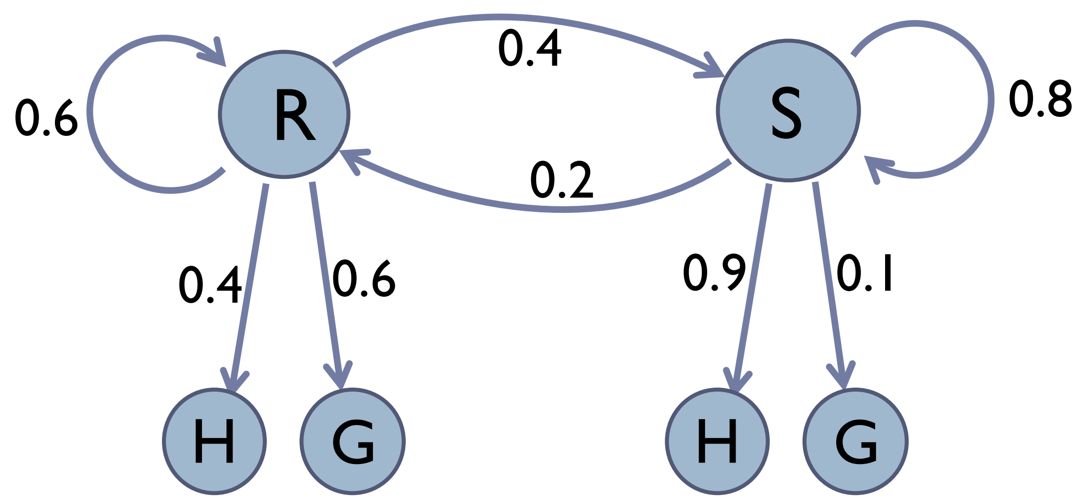
- Lets go back to the Rainy/Sunny model and assume one can be Happy(H) or Grumpy(G) accord. to the weather.
- Probabilities from the hidden states (R,S) to the observable states (H,G) are called emission probabilities.
- Starting probabilities: $P(R_0)=0.5$ and $P(S_0)=0.5$



Hidden Markov Model Example

- $P(R_1|H_1)$=? 16/70 = 0.229
- Using Bayes rule: $=P(H_1|R_1)\ P(R_1)/P(H_1)$

  where $P(R_1)= P(R_1|R_0)\ P(R_0)+P(R_1|S_0)\ P(S_0)$
- We just estimated a hidden variable based on an observed variable.



Forward-Backward Algorithm for HMMs

- We are interested in finding the conditional distribution over a hidden variable at time n, given the values of all the observable variables:

$$P(z_n|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|z_n)P(z_n)}{P(\boldsymbol{X})}$$

$$= \frac{P(x_1, x_2, \dots, x_n|z_n)P(x_{n+1}, \dots, x_N|z_n)P(z_n)}{P(\boldsymbol{X})}$$

$$= \frac{P(x_1, x_2, \dots, x_n, z_n)P(x_{n+1}, \dots, x_N|z_n)}{P(\boldsymbol{X})} = \frac{\alpha(z_n)\beta(z_n)}{P(\boldsymbol{X})}$$

Viterbi Algorithm for HMMs

- We might also want to estimate the most likely sequence of all the hidden variables given a sequence of observations.
- We need to maximize the joint distribution over $\boldsymbol{Z}$

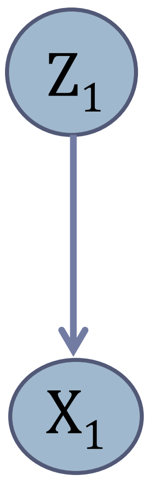
$$\max_{\boldsymbol{Z}} P(\boldsymbol{X}, \boldsymbol{Z}) = \max_{\boldsymbol{Z}} P(z_1) \left[\prod_{n=2}^{N} P(z_n | z_{n-1}) \right] \prod_{n=1}^{N} P(x_n | z_n)$$

- Alternatively we can maximize the logarithm of these expressions so that we do not have to multiply many small numbers.

Measurement – Prediction Cycle

$$P(Z_1|X_1)=P(X_1|Z_1)\cdot P(Z_1)/P(X_1)$$

Measurement update



$$P(Z_2) = \sum_{Z_1} P(Z_1) \cdot P(Z_2|Z_1)$$

Prediction update

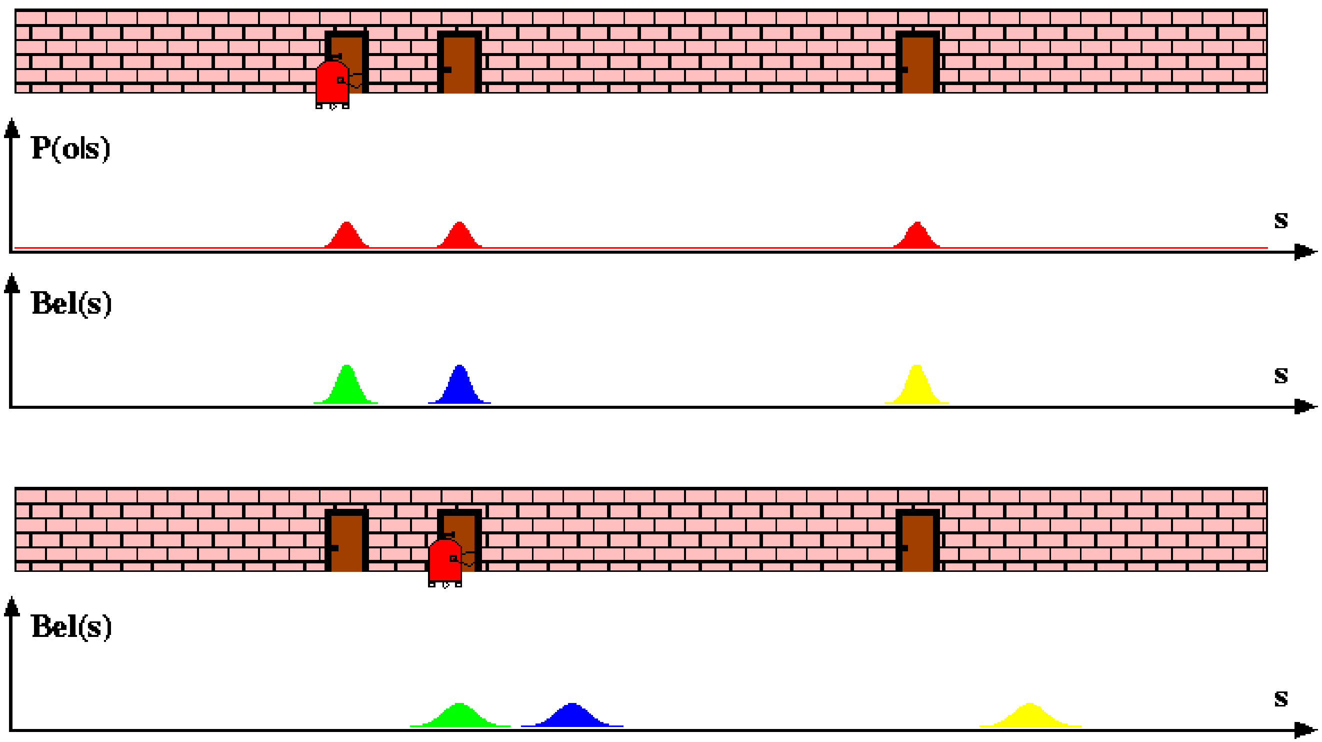
Algorithms that use this cycle are also called filters.
E.g. Kalman filter, particle filter.

Robot Localization Example with Measurement – Prediction update cycle

Senses the environment

Measurement update (Bayes rule) (priors were same)

Prediction update (bumps are shifted and flattened due to convolution)

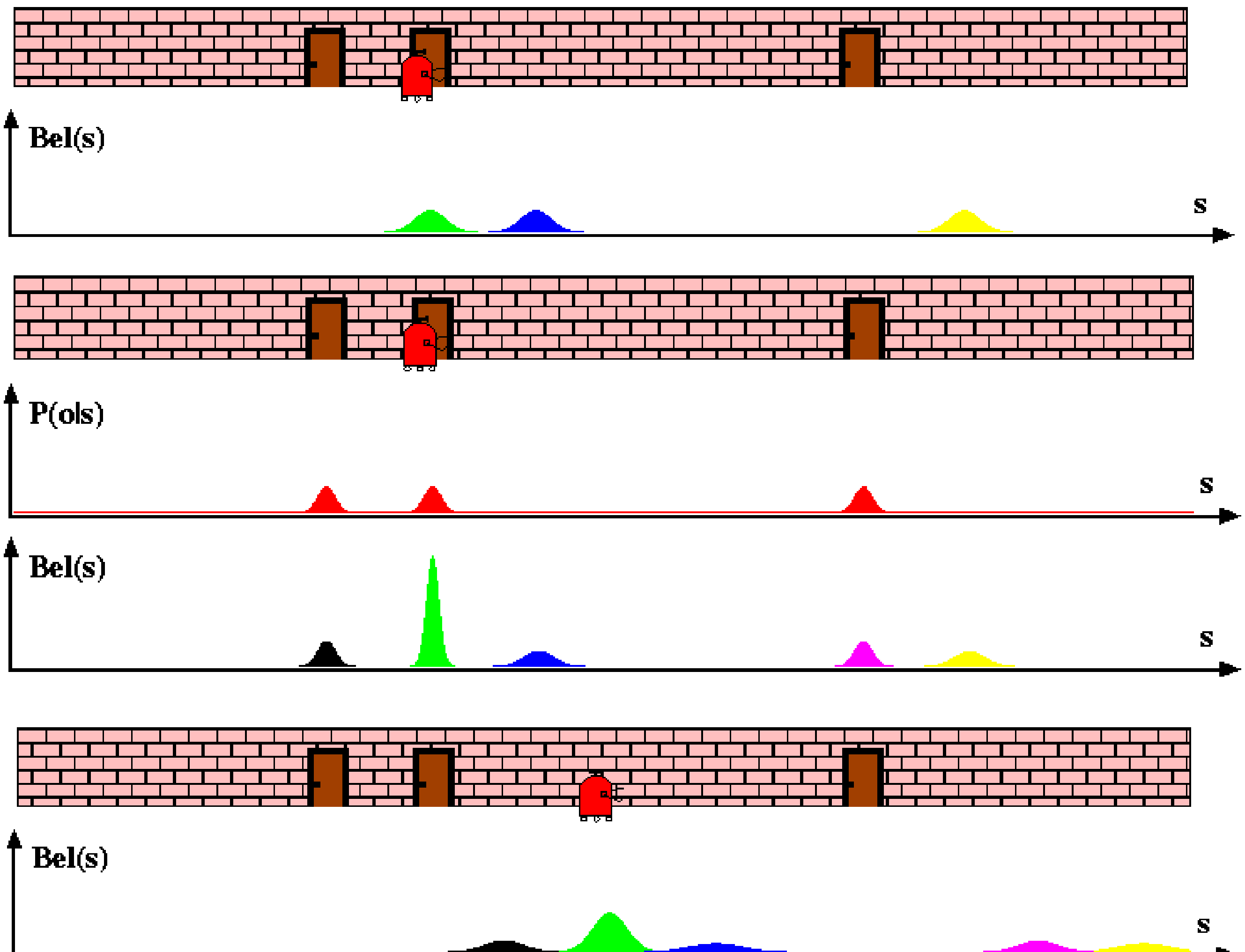


Robot Localization Example with Measurement – Prediction update cycle

Senses the environment

Measurement update (priors were different)

Another prediction update



Slides from *robots.stanford.edu/probabilistic-robotics/ppt/kalman.ppt*