Intelligent Robotics

Probabilistic State Estimation

Mohan Sridharan
School of Computer Science
University of Birmingham, UK

October 7, 2019

Lecture Outline

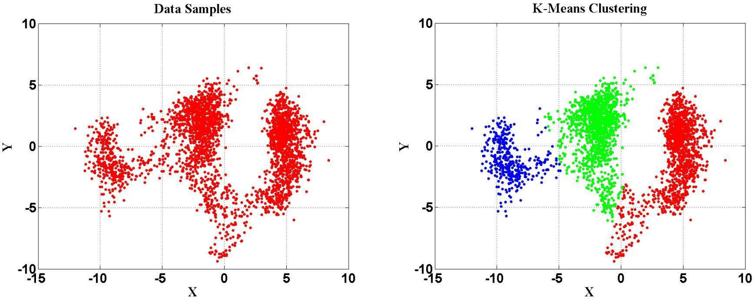
- Bayesian classification.
- Bayesian inference.

Classification Basics

- Broad categories: supervised (labeled samples); unsupervised (no labeled samples).
- Group data based on similarity measures.
- Many sophisticated methods:
  - Supervised: decision trees, support vector machines, neural networks.
  - Unsupervised: nearest neighbors, clustering.
- Choice of classifier based on data and application.
- *Probabilistic methods model the noise in input data!*

Clustering Data Samples

- K-Means clustering of input data samples.
- Data grouped into three clusters.



Bayesian Classification

- Bayes' rule (once again):

$$p(x, y) = p(x|y) \cdot p(y) = p(y|x) \cdot p(x)$$

$$p(x|y) = \frac{p(y|x) \cdot p(x)}{p(y)} = \frac{\text{likelihood . prior}}{\text{normalizer}}$$

- Classify based on Bayes *decision rule*:

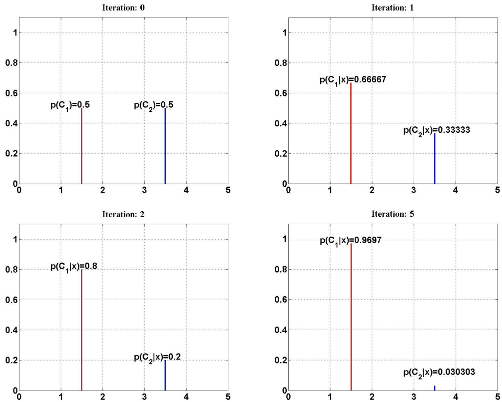
$$p(C_1|x) > p(C_2|x) \implies \text{choose } C_1\text{; else choose } C_2$$

- Decision rule extends to multiple classes:

$$p(C_i|x) > p(C_j|x) \;\; \forall j \neq i \implies \text{choose } C_i$$

Illustrative Example 1

- C_1 : $room_1$; C_2 : $room_2$; x : $data$ (e.g., specific door).
- $p(C_1) = p(C_2) = 0.5$; $p(x|C_1) = 0.6$; $p(x|C_2) = 0.3$



Multi-Class Extension

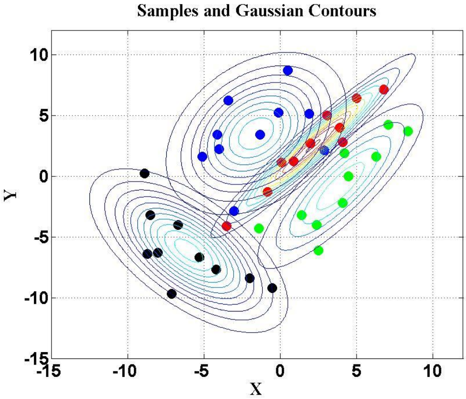
- Model *likelihoods* and *priors* based on training samples.
- Update belief incrementally based on evidence.
- Use multi-class Decision rule:

$$p(C_i|x) > p(C_j|x) \;\; \forall j \neq i \implies \text{choose } C_i$$

- **Question**: representation to use for likelihoods?
- **Answer**: use functions with well-understood properties, e.g., **Gaussians**.

Illustrative Example 2

- Four-class problem; ten training data samples per class.
- Model individual class likelihoods as Gaussians.



Illustrative Example 2: Modeling

- Compute Gaussian means and covariances:

$$\mu_1 = [2.16, 2.49]; \quad \mu_2 = [3.95, -0.84]$$
$$\mu_3 = [-1.57, 3.5]; \quad \mu_4 = [-6, -6.14]$$

$$\Sigma_1 = \begin{pmatrix} 9.32 & 10.12 \\ 10.12 & 11.85 \end{pmatrix}$$

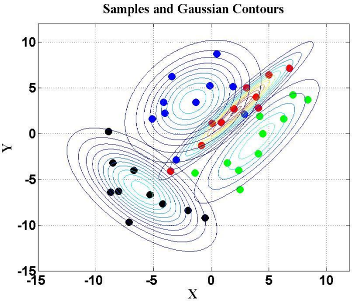
$$\Sigma_2 = \begin{pmatrix} 8.36 & 8.87 \\ 8.87 & 13.02 \end{pmatrix}$$

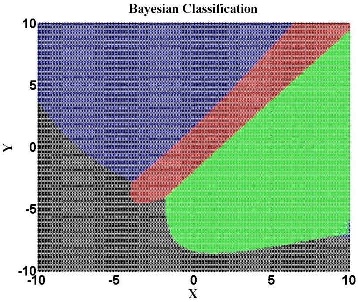
$$\Sigma_3 = \begin{pmatrix} 7.63 & 2.98 \\ 2.98 & 9.78 \end{pmatrix}$$

$$\Sigma_4 = \begin{pmatrix} 8.62 & -5.71 \\ -5.71 & 9.26 \end{pmatrix}$$

Illustrative Example 2: Classification

- Decision boundaries for all four classes:





Summary

- Elegant belief update and decision rule for classification.
- Little or no tuning of arbitrary thresholds.

- Bayes error: *minimum classification error that cannot be eliminated.*

- Challenge 1: *what function and parameters to use for modeling likelihoods and priors?*

- Challenge 2: *how to obtain enough data to model the likelihoods and priors?*

For more information

- C. Bishop. *Pattern Recognition and Machine Learning*. Springer publishing house, 2007.
- R. Duda and P. Hart and D. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
- D. Stork and E. Yom-Tov. *Computer Manual in MATLAB to accompany Pattern Classification*. Wiley-Interscience, 2004.
- *Weka 3: Data Mining Software in Java*, 2010. http://www.cs.waikato.ac.nz/ml/weka/.
- *MATLAB Statistics Toolbox*. http://www.mathworks.com/products/statistics/

Lecture Outline

- Bayesian classification.
- Bayesian inference.

The Framework

- Inputs:
  - Stream of observations z and actions u: $\{u_1, z_1, \ldots, u_t, z_t\}$
  - Sensor model: $p(z|x)$
  - Action model: $p(x'|u, x)$
  - Prior probability of system state: $p(x)$
- Outputs:
  - Estimate the state $\boldsymbol{x}$ of a *dynamical system*.
  - Posterior of state, called the belief:

$$bel(x_t) = p(x_t|u_1, z_1, \ldots, u_t, z_t)$$

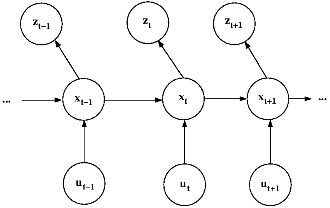
Markov Assumption

- First-order Markov (conditional independence) assumption:

$$p(x_t|x_0, \ldots, x_{t-1}) = p(x_t|x_{t-1})$$

- Bayesian filtering:

$$p(z_t|x_{0:t}, z_{1:t}, u_{1:t}) = p(z_t|x_t)$$
$$p(x_t|x_{1:t-1}, z_{1:t}, u_{1:t}) = p(x_t|x_{t-1}, u_t)$$



Bayes Filters 1

- Bayes rule:

$$\begin{aligned} bel(x_t) &= p(x_t | u_{1:t}, z_{1:t}) \\ &\propto p(z_t | x_t, u_1, z_1, \ldots, u_t)\ p(x_t | u_1, z_1, \ldots, u_t) \end{aligned}$$

- Markov assumption:

$$\begin{aligned} bel(x_t) &\propto p(z_t | x_t, u_1, z_1, \ldots, u_t)\ p(x_t | u_1, z_1, \ldots, u_t) \\ &= p(z_t | x_t)\ p(x_t | u_1, z_1, \ldots, u_t) \end{aligned}$$

Bayes Filters 2

- Probability expansion:

$$bel(x_t) \propto p(z_t|x_t)\, p(x_t|u_1, z_1, \ldots, u_t)$$
$$= p(z_t|x_t) \int p(x_t|u_{1:t}, z_{1:t-1}, x_{t-1})p(x_{t-1}|u_{1:t}, z_{1:t-1})\, dx_{t-1}$$

- Markov assumption:

$$bel(x_t) \propto p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, p(x_{t-1}|u_1, z_1, \ldots, u_t)\, dx_{t-1}$$

Bayes Filters 3

- Markov assumption:

$$bel(x_t) \propto p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, p(x_{t-1}|u_1, z_1, \ldots, u_t)\, dx_{t-1}$$
$$= p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, p(x_{t-1}|u_1, z_1, \ldots, z_{t-1})\, dx_{t-1}$$

- Recursion:

$$bel(x_t) = \eta\ p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, bel(x_{t-1})\, dx_{t-1}$$

Bayes Filters Summary

- Recursive belief update based on Markov assumption:

$$
\begin{aligned}
bel(x_t) &= p(x_t|u_{1:t}, z_{1:t}) \\
&\propto p(z_t|x_t, u_1, z_1, \ldots, u_t)\, p(x_t|u_1, z_1, \ldots, u_t) \\
&= p(z_t|x_t)\, p(x_t|u_1, z_1, \ldots, u_t) \\
&= p(z_t|x_t) \int p(x_t|u_{1:t}, z_{1:t-1}, x_{t-1}) p(x_{t-1}|u_{1:t}, z_{1:t-1})\, dx_{t-1} \\
&= p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, p(x_{t-1}|u_1, z_1, \ldots, u_t)\, dx_{t-1} \\
&= p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, p(x_{t-1}|u_1, z_1, \ldots, z_{t-1})\, dx_{t-1}
\end{aligned}
$$

$$
bel(x_t) = \eta\; p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\, bel(x_{t-1})\, dx_{t-1}
$$

Bayes Inference

- Bayes prediction and correction:

$$\forall x_t : \; bel(x_t) = \eta \; p(z_t|x_t) \int p(x_t|u_t, x_{t-1}) \, bel(x_{t-1}) \, dx_{t-1}$$

$$\forall k : \; p_{k,t} = \eta \; p(z_t|X_t = x_k) \sum_i p(X_t = x_k|u_t, X_{t-1} = x_i) \, p_{i,t-1}$$

- Bayes filter:

$$\forall x_t : \overline{bel}(x_t) = \int p(x_t|u_t, x_{t-1}) \, bel(x_{t-1}) \, dx_{t-1}$$

$$bel(x_t) = \eta \; p(z_t|x_t) \, \overline{bel}(x_t)$$

- Discrete Bayes filter:

$$\forall k : \overline{p}_{k,j} = \sum_i p(X_t = x_k|u_t, X_{t-1} = x_i) \, p_{i,t-1}$$

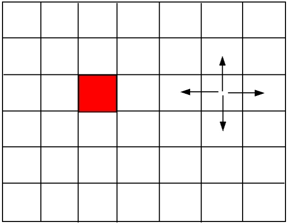
$$p_{k,j} = \eta \; p(z_t|X_t = x_k) \, \overline{p}_{k,j}$$

Examples

- Pictorial representation of discrete Bayes:

$$\forall k : \overline{p}_{k,j} = \sum_i p(X_t = x_k | u_t, X_{t-1} = x_i)\, p_{i,t-1}$$

$$p_{k,j} = \eta\; p(z_t | X_t = x_k)\, \overline{p}_{k,j}$$



- Many instances: Kalman filters, Particle filters, Bayesian Networks, Partially Observable Markov Decision Processes (POMDPs), Hidden Markov Models (HMMs)...

Summary

- Bayesian inference is a general framework for probabilistic state estimation.
- Markov assumption, although not always true, allows for elegant belief updates.
- Incorporates changes in system dynamics independent of the observations of the system.
- **Applications:** *computer vision, robotics, agricultural estimation, climate informatics, and many more...*.