Perception and Bayes Rule

CS 3630

Perception

Perception is the process of inferring the state of the world (and possibly of the robot itself) using sensor measurements and other contextual information.

Sensing vs perception

- Sensor models are *forward* models.
  - Given a description of the world and a model of the sensor,
  - ➢Determine the conditional probability mapping

$$P(SensorReading \mid State)$$

- Perception is concerned with the *inverse* problem.
  - Given a set of sensor readings and (possibly extra contextual information),
  - ➢Infer the probability map associated to the world state

$$P(State \mid SensorReadings, Context)$$

  - Context could include previous sensor readings, knowledge about the robot's actions, etc.

Bayes theorem

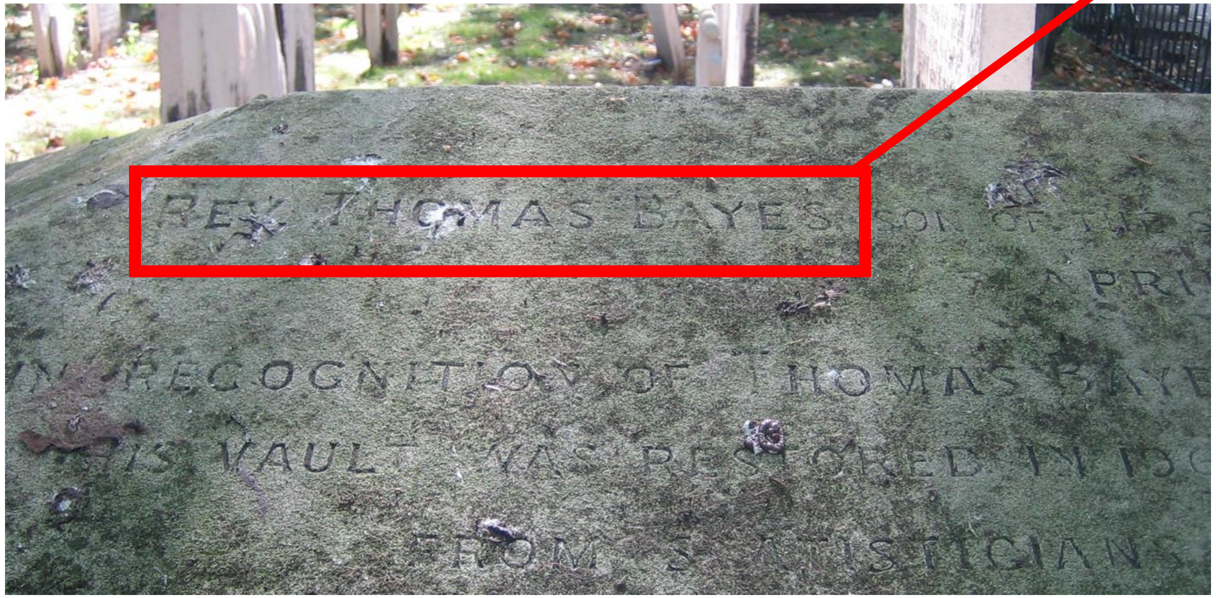
We want to compute:

$$P(State \mid SensorReadings, Context)$$

But we are given

$$P(SensorReading \mid State)$$

Bayes derived his famous inversion equation for just this purpose.



Bayes is probably buried here
(Bunhill Fields Cemetery, London).

Bayes Theorem

We know that conjunction is commutative:

$$P(A,B) = P(B,A)$$

Using the definition of conditional probability:

$$P(B|A)P(A) = P(B,A) = P(A,B) = P(A|B)P(B)$$

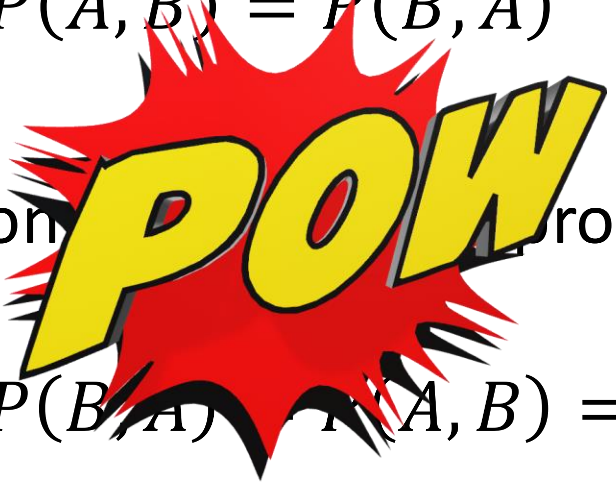
$$P(B|A)P(A) = P(A|B)P(B)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Bayes Theorem

We know that conjunction is commutative:

$$P(A,B) = P(B,A)$$



Using the definition … probability:

$$P(B|A)P(A) = P(B,A) = P(A,B) = P$$

$$|A)P(A) = P(A|B)P(B)$$

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Example

We roll one die, and an observer tells us things about the outcome. We want to know if $X = 4$.

- Before we know anything, we believe $P(X = 4) = \frac{1}{6}$. PRIOR
- Now, suppose the observer tells us that X is even. EVIDENCE

$$P(X = 4 \mid X \text{ even}) = \frac{P(X \text{ even} \mid X = 4)P(X = 4)}{P(X \text{ even})} = \frac{1 \times \frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$ Bayes

- We could also use Bayes to infer $P(X = \text{even} \mid X = 4)$:

$$P(X \text{ even} \mid X = 4) = \frac{P(X = 4 \mid X \text{ even})P(X \text{ even})}{P(X = 4)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{6}} = 1$$ Somewhat less interesting

Interpreting Bayes theorem

- The individual terms on the right-hand side have intuitive interpretations
- We observe y, and we want to update our belief about x based on this observation.
- In this case,
  - We can think of y as evidence and $P(y)$ is the probability of observing this particular evidence.
  - The conditional probability $P(y|x)$ is called the likelihood of the evidence (given x).
  - The probability $P(x)$ is the prior probability for x.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \qquad posterior = \frac{\text{likelihood } \cdot \text{prior}}{\text{evidence}}$$

Example

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \qquad posterior = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

x is robot pose and y is sensor data

$p(x)$	→	*Prior* probability distribution
$p(y\|x)$	→	*Likelihood*, sensor model that we've seen before
$p(y)$	→	*Evidence*, does not depend on x

Given or observed.

$p(x\|y)$	→	*Posterior* (conditional) probability distribution

Perception.

About likelihoods…

Why do we call the conditional probability $p(y|x)$ a *likelihood*, but we call $p(x|y)$ the *posterior*?

- Likelihood represents $p(data|parameters)$
  - The probability of observing your data given specific parameter values, in this case some object or robot state
  - Not a probability distribution
- Posterior $p(parameters|data)$
  - The probability of parameter values being true given your observed data
  - Is a proper probability distribution

Example

- For our conductivity sensor, we defined the conditional probabilities $p(x|C)$ for each category C.
- The rows in this table represent conditional probabilities of sensor readings given object category.
- The columns in this table represent the likelihood of each category for a given sensor measurement.

Category (C)	P(False|C)	P(True|C)	Conditional probabilities – they sum to one!
Cardboard	0.99	0.01	$p(x\|Cardboard)$
Paper	0.99	0.01	$p(x\|Paper)$
Cans	0.1	0.9	$p(x\|Cans)$
Scrap Metal	0.15	0.85	$p(x\|Metal)$
Bottle	0.95	0.05	$p(x\|Bottle)$
	$p(False\|C)$	$p(True\|C)$	

Likelihoods of categories – they ***do not*** sum to one!

Normalization Coefficient

$$P(x|z) = \frac{P(z|x)P(x)}{P(z)}$$

Note that the denominator is independent of x, and as a result will typically be the same for any value of x in the posterior $P(x|z)$.

Therefore, we typically represent the normalization term by the coefficient

$$\eta = [P(z)]^{-1}$$

and Bayes equation is written as

$$P(x|z) = \eta P(z|x)P(x)$$

Posterior

Normalization Constant

Likelihood

Prior

Marginal Distributions

- Suppose we are given the joint probability map $P(x, y)$.
- Can we compute $P(y)$?
- Simply sum the probabilities for every joint event in which x occurs with some outcome y_i:

$$P(y) = \sum_{x_i} P(x_i, y)$$

We've seen this in an earlier example:

- Roll two dice and denote by x_1 and x_2 the number of dots showing on their faces.
- What is the probability that $\mathrm{x}_1 = 6$?
- We sum the probabilities for joint events in which $\mathrm{x}_1 = 6$: {(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

$$P(x_1 = 6) = \sum_{y \in \{1,\ldots,6\}} P(x_1 = 6, y) = \frac{1}{6}$$

This is called **marginalization**, and $P(x_1 = 6)$ is called a **marginal probability**.

Law of total probability

- We are typically not given the joint distribution $P(x, y)$.
- However, we are often given the conditional distribution $P(y|x)$.
- We can use this to compute the marginal $P(y)$:

$$P(y) = \sum_{x_i} P(x_i, y)$$

Law of total probability

- We are typically not given the joint distribution $P(x, y)$.
- However, we are often given the conditional distribution $P(y|x)$.
- We can use this to compute the marginal $P(y)$:

$$P(y) = \sum_{x_i} P(x_i, y) = \sum_{x_i} P(y|x_i)P(x_i)$$

Sum over all the ways in which y can occur (i.e., all possible values x_i that can occur with y).

The probability that y occurs, given x_i

Multiplied by the prior probability of x_i

Example

- We can use the law of total probability to compute the normalization constant in Bayes equation.
- For example, to compute the probability that the conductivity sensor will return the value $True$, let $y = True$, and sum over the five categories:

$$P(True) = \sum_{C_i} P(True|C_i)P(C_i)$$

$$= P(True|Cardboard)P(Cardboard) + P(True|Paper)P(Paper) + P(True|Can)P(Can) + P(True|Scrap\ Metal)P(Scrap\ Metal) + P(True|Bottle)P(Bottle)$$

Causal vs. Diagnostic Reasoning

- $P(Paper|z)$ is diagnostic.
- $P(z|Paper)$ is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

Comes from sensor model.

Comes from experience.

$$P(Paper|z) = \frac{P(z|Paper)P(Paper)}{P(z)}$$

Use law of total probability: $P(z) = \sum_y P(z|y)P(y)$

Bayes law, one last time

We can expand the denominator using the law of total probability to obtain:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}$$

Again, note that the denominator does not depend on x.