KRR with Uncertainty

Review of Probability

Ira A. Fulton Schools of Engineering
Arizona State University

Objectives

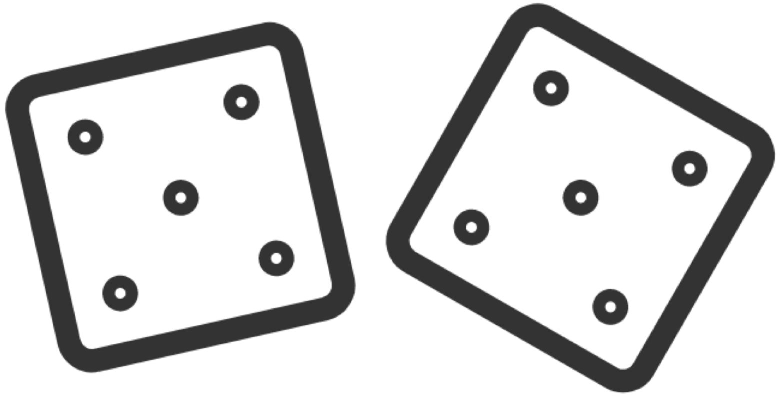
Objective

Review the basics of probability theory

Probability Theory

- **We identify a proposition with an actual event represented by the proposition**

- **Basic axioms of probability**
  - $0 \leq P(A) \leq 1$ for all propositions A
  - $P(tautology) = 1$
  - $P(A \vee B) = P(A) + P(B)$ if A and B are "mutually exclusive"



Random Variables and Possible Worlds (1 of 3)

- **A (discrete) random variable is similar to (multi-valued) atoms in propositional logic, but the agent is uncertain about its value**

- **A random variable X is associated with the domain, denoted $dom(X)$, the set of values X can take.**
  - Loc is a random variable whose domain is the set of locations
  - $Smoking$ is a random variable whose domain is Boolean.

Random Variables and Possible Worlds (2 of 3)

- A **possible world *w*** specifies an assignment to each random variable.

- **Example:** We model only 2 Boolean variables $Smoking$ and $Cancer$. Then there are $2^2=4$ distinct possible worlds:

$$w_1\text{:}\quad Smoking = T \ \wedge\ Cancer = T$$

$$w_2\text{:}\quad Smoking = T \ \wedge\ Cancer = F$$

$$w_3\text{:}\quad Smoking = F \ \wedge\ Cancer = T$$

$$w_4\text{:}\quad Smoking = F \ \wedge\ Cancer = F$$

Random Variables and Possible Worlds (3 of 3)

- A **possible world** can be identified with an interpretation in a logical language

- $w \vDash X = x$ means variable X is assigned value x in world w

- Define a **nonnegative measure** $\mu(w)$ to possible worlds w such that the measures of the possible worlds **sum to 1**

- The **probability of proposition** f is defined by:

$$p(f) = \sum_{w \vDash f} \mu(w)$$

Joint Distribution

- The **joint distribution** over random variables $X_1, \ldots, X_n$:
  - a probability distribution over the joint random variable $< X_1, \ldots, X_n >$ with domain $dom(X_1) \times \ldots \times dom(X_n)$
  - analogous to truth tables in logical language

Weather	Temperature	$\mu\,(w)$
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

- **In general, each row corresponds to an assignment** $X_1 = x_1, \ldots, X_n = xn$ **and its probability** $P(X_1 = x_1, \ldots, X_n = xn)$

Marginalization (1 of 2)

Given the joint distribution, we can compute distributions over smaller sets of variables through **marginalization**:

$$\boxed{P(X = x) = \Sigma_{z \in \mathrm{dom}(Z)} P(X = x, Z = z)}$$

- We also write this as $P(X) = \Sigma_{z \in dom(Z)} P(X, Z = z)$.

This corresponds to summing out a dimension in the table.

Marginalization (2 of 2)

- **The new table still sums to 1. It must, since it's a probability distribution!**

Weather	Temperature	$\mu\,(w)$
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Temperature	$\cancel{\mu\,(w)}$ P(T)
hot	0.15
mild	0.55
cold	0.3

In general $P(A) \;=\; \Sigma_{w \vDash A}\, \mu(w)$

Conditioning (1 of 3)

- Conditioning specifies how to revise beliefs based on new information.

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

Possible world	Weather	Temperature	$\mu(w)$
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~ 0
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~ 0
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~ 0

T	$P(T\|W=sunny)$
hot	
mild	
cold	

Conditioning (2 of 3)

- Conditioning specifies how to revise beliefs based on new information.

- You have a prior for the joint distribution of weather and temperature, and the marginal distribution of temperature

Possible world	Weather	Temperature	$\mu(w)$
w_1	sunny	hot	0.10
w_2	sunny	mild	0.20
w_3	sunny	cold	0.10
~~w_4~~	~~cloudy~~	~~hot~~	~~0.05~~
~~w_5~~	~~cloudy~~	~~mild~~	~~0.35~~
~~w_6~~	~~cloudy~~	~~cold~~	~~0.20~~

T	*P (T|W=sunny)*
hot	0.10/0.40=0.25
mild	0.20/0.40=0.50
cold	0.10/0.40=0.25

Conditioning (3 of 3)

- Conditioning specifies how to revise beliefs based on new information.

- Evidence e rules out possible worlds incompatible with *e*

$$\mu_e(w) = \begin{cases} \dfrac{1}{P(e)} \times \mu(w) & \text{if} \quad w \models e \\ 0 & \text{if} \quad w \not\models e \end{cases}$$

Bayes rule

- By definition, we know that $P(h|e) = \dfrac{P(h \wedge e)}{P(e)}$

- We can rearrange terms to show: $P(h \wedge e) = P(h|e) \cdot P(e)$

- Similarly, we can show: $P(e \wedge h) = P(e|h) \cdot P(h)$

- Since $e \wedge h$ and $h \wedge e$ are identical, we have:

$$P(h|e) = \frac{P(e|h) \cdot P(h)}{P(e)}$$

Example of Bayes rule

- **On average, the alarm rings once a year**
  - $P(alarm) = 1/365$
- **If there is a fire, the alarm will almost always ring**
  - $P(alarm|fire) = 0.999$
- **On average, we have a fire every 10 years**
  - $P(fire) = 1/3650$
- **Q. The fire alarm rings. What is the probability there is a fire?**

$$P(fire \mid alarm) = \frac{P(alarm \mid fire) \times P(fire)}{P(alarm)} = \frac{0.999 \times \frac{1}{3650}}{\frac{1}{365}} = 0.0999$$

Product Rule

We know

- $P(f_1 \wedge f_2) = P(f_1) \times P(f_2 | f_1)$
- $P(f_1 \wedge f_2 \wedge f_3) = P(f_1) \times P(f_2 | f_1) \times P(f_3 | f_1, f_2)$

In general,

- $P(f_1 \wedge f_2 \wedge \cdots f_n) = P(f_1) \times P(f_2 | f_1) \times P(f_3 | f_1, f_2) \times \cdots \times P(f_n | f_1, \ldots f_{n-1})$

Marginal Independence

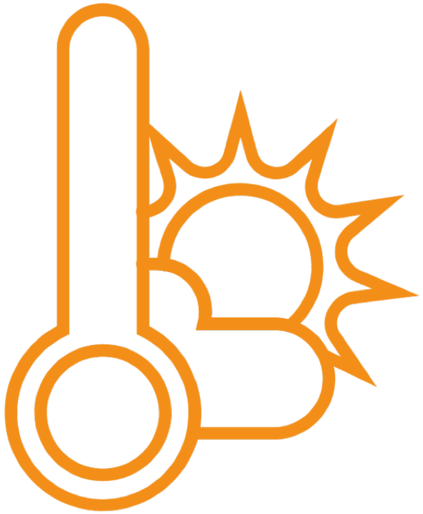
Definition (Marginal Independence)

Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j, y_k \in dom(Y)$, the following equation holds:

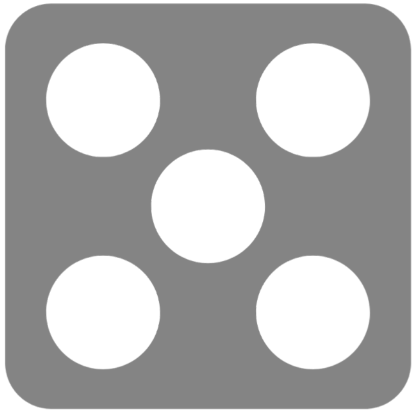
$$
\begin{aligned}
&P(X = x_i \mid Y = y_j) \\
&\quad = P(X = x_i \mid Y = y_k) \\
&\quad = P(X = x_i)
\end{aligned}
$$

Marginal Independence, cont'd

- **Intuitively: If X and Y are marginally independent, then**
  - learning $Y = y$ does not change your belief in X
  - and this is true for all values y that Y could take.



- **For example, weather is marginally independent from the result of a die throw.**



Example: Marginal Independence (1 of 2)

Definition (Marginal Independence)

Random variable X is (marginally) independent of random variable Y if, for all $x_i \in dom(X)$, $y_j, y_k \in dom(Y)$, the following equation holds:

$$P(X = x_i \mid Y = y_j) = P(X = x_i \mid Y = y_k) = P(X = x_i)$$

Q: Are A and B marginally independent?

yes

no

$P(B=h) = 0.15$

$P(B=h \mid A=s) = \frac{P(B=h \wedge A=s)}{P(A=s)} = \frac{0.1}{0.4} = 0.25$

$P(B=h \mid A=c) = \frac{P(B=h \wedge A=c)}{P(A=c)} = \frac{0.05}{0.6} = 0.083$

A	B	P(A,B)
sunny	hot	0.10
sunny	mild	0.20
sunny	cold	0.10
cloudy	hot	0.05
cloudy	mild	0.35
cloudy	cold	0.20

Example: Marginal Independence (2 of 2)

Definition (Marginal Independence)

Random variable X is (marginally) independent of random variable Y if, for all $x_i \in \mathrm{dom}(X)$, $y_j \in \mathrm{dom}(Y)$ and $y_k \in \mathrm{dom}(Y)$, the following equation holds:

$$P(X = x_i \mid Y = y_j)$$
$$= P(X = x_i \mid Y = y_k)$$
$$= P(X = \cancel{xi})$$

x_i

Q: Are C_1 and C_2 marginally independent?

$P(C_1 = h) = 0.5$

$P(C_1 = h \mid C_2 = h) = \frac{P(C_1 = h \wedge C_2 = h)}{P(C_2 = h)} = \frac{0.25}{0.5} = 0.5$

C_1	C_2	$P(C_1,C_2)$
head	head	0.25
head	tail	0.25
tail	head	0.25
tail	tail	0.25

Conditional Independence (1 of 2)

Definition (Conditional Independence)

Random variable X is (conditionally) independent of random variable Y given random variable Z if, for all $x_i \in dom(X)$, $y_j, y_k \in dom(Y)$ and $z_m \in dom(Z)$, the following equation holds:

$$P(X = x_i \mid Y = y_j, Z = z_m)$$

$$= P(X = x_i \mid Y = y_k, Z = z_m)$$

$$= P(X = x_i \mid Z = z_m)$$

Conditional Independence (2 of 2)

- **Intuitively: If X and Y are conditionally independent given Z, then**
  - Learning that $Y = y$ does not change your belief in X when we already know $Z = z$
  - and this is true for all values y that Y could take and all values z that Z could take.

Questions

- Q: If X and Y are conditionally independent given Z, are they marginally independent?

- Q: If X and Y are marginally independent, are they conditionally independent given Z?

Conditional vs. Marginal Independence

Two variables can be

- Both marginally ~~nor~~ and conditionally independent
  - PhoenixSunsWins and LightOn
  - PhoenixSunsWins and LightOn given PowerOn

- Neither marginally nor conditionally independent
  - Temperature and Cloudiness
  - Temperature and Cloudiness given Wind

- Conditionally but not marginally independent
  - ExamGrade and AssignmentGrade
  - ExamGrade and AssignmentGrade given UnderstoodMaterial

- Marginally but not conditionally independent
  - SmokingAtSensor and Fire
  - SmokingAtSensor and Fire given Alarm

Exploiting Conditional Independence

Recall the chain rule

Theorem (Chain Rule)

$$P(f_n \wedge \cdots \wedge f_1) = \prod_{t=1}^{n} P(f_i \mid f_{i-1} \wedge \cdots \wedge f_1)$$

- e.g., $P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|A,B,C)$

$15 = 1 + 2 + 4 + 8$

If D is conditionally independent of A and B given C,

$$P(A,B,C,D) = P(A) \times P(B|A) \times P(C|A,B) \times P(D|C)$$

$9 = 1 + 2 + 4 + 2$

Under independence, we gain **compactness**.

Wrap-Up