Chapter 1
Probabilistic Reasoning

- Intelligence
  - Understanding, reasoning, planning, problem-solving, …
  - Natural intelligence / Artificial intelligence
- Artificial intelligence(AI)
  - Intelligence demonstrated by machines
- Machine learning (ML)
  - A part of artificial intelligence
  - Study of computer algorithms that improve automatically through experience and by the use of data
  - Goal - understanding and prediction

- Machine learning
    - Supervised learning / Unsupervised learning / Reinforcement learning
    - Statistical learning / Artificial neural networks
- Statistical learning
    - A framework for machine learning
    - Drawing from the fields of statistics and functional analysis

1.1 Probability Refresher

- Joint distribution

$$p(x, y) = p(x \text{ and } y)$$

- Marginalisation

$$p(x) = \sum_{y} p(x, y)$$

$$p(x_1, \dots, x_{i-1}, x_{i+1}, \dots x_n) = \sum_{x_i} p(x_1, \dots, x_n)$$

- Conditional Probability / Bayes' Rule

$$p(x|y) \equiv \frac{p(x,y)}{p(y)}$$

- Since $p(x,y) = p(y,x)$, we have

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

or equivalently,

$$\begin{aligned} p(x,y) &= p(x|y)p(y) \\ &= p(y|x)p(x) \end{aligned}$$

- Probability Density Function

$$f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- Probability that $x \in [a, b]$

$$p(a \leq x \leq b) = \int_{a}^{b} f(x)dx$$

1.1.1 Interpreting Conditional Probability

- Dart

$$p(region\ i) = \frac{1}{20}$$

- Example
  - the probability that Randy has hit the region 5
  - conditioned on the information that he hasn't hit the region 20

$$p(region\ 5 \mid not\ region\ 20)$$

- Independence
  - Variables x and y are independent if knowing the state of one variable gives no extra information about the other variable
  - In equation

$$p(x, y) = p(x)p(y)$$

    or equivalently

$$p(x|y) = p(x)$$

- Example

  $$p(x = \mathrm{a},\ \ y = 1) = 1,\ p(x = \mathrm{a},\ \ y = 2) = 0,$$
  $$p(x = \mathrm{b},\ \ y = 2) = 0,\ p(x = \mathrm{b},\ \ y = 1) = 0$$

  Are x and y dependent?

- Note

  variable x, value a

- Conditional Independence
  - Sets of variables $\mathcal{X}$ and $\mathcal{Y}$ are independent under the condition $\mathcal{Z}$
$$\mathcal{X} \coprod \mathcal{Y} | \mathcal{Z}$$
that is,
$$p(\mathcal{X}, \mathcal{Y} | \mathcal{Z}) = p(\mathcal{X} | \mathcal{Z}) p(\mathcal{Y} | \mathcal{Z})$$
  - Dependent
$$\mathcal{X} \prod \mathcal{Y} | \mathcal{Z}$$
  - Note
$$\mathcal{X} \coprod \mathcal{Y} = \mathcal{X} \coprod \mathcal{Y} | \emptyset$$
$$\mathcal{X} \prod \mathcal{Y} = \mathcal{X} \prod \mathcal{Y} | \emptyset$$

- Independence implications

$$a \coprod b,\ b \coprod c \ \Rightarrow\ a \coprod c$$

- Prove that

$$p(a,b) = p(a)p(b),\ p(b,c) = p(b)p(c) \Rightarrow p(a,c) = p(a)p(c)$$

1.1.2 Probability Tables

- Populations of UK
  - England(E) – 60776238
  - Scotland(S) – 5116900
  - Wales(W) – 2980700

$$\begin{pmatrix} p(Cnt = \mathrm{E}) \\ p(Cnt = \mathrm{S}) \\ p(Cnt = \mathrm{W}) \end{pmatrix} = \begin{pmatrix} .88 \\ .08 \\ .04 \end{pmatrix}$$

- Mother Tongue(MT)

$$\begin{array}{lll}
p(MT = \text{Eng}|Cnt = \text{E}) = 0.95 & p(MT = \text{Eng}|Cnt = \text{S}) = 0.7 & p(MT = \text{Eng}|Cnt = \text{W}) = 0.6 \\
p(MT = \text{Scot}|Cnt = \text{E}) = 0.04 & p(MT = \text{Scot}|Cnt = \text{S}) = 0.3 & p(MT = \text{Scot}|Cnt = \text{W}) = 0.0 \\
p(MT = \text{Wel}|Cnt = \text{E}) = 0.01 & p(MT = \text{Wel}|Cnt = \text{S}) = 0.0 & p(MT = \text{Wel}|Cnt = \text{W}) = 0.4
\end{array}$$

- $p(Cnt, MT) = p(MT|Cnt)p(Cnt)$

$$\begin{pmatrix} 0.95 \times 0.88 & 0.7 \times 0.08 & 0.6 \times 0.04 \\ 0.04 \times 0.88 & 0.3 \times 0.08 & 0.0 \times 0.04 \\ 0.01 \times 0.88 & 0.0 \times 0.08 & 0.4 \times 0.04 \end{pmatrix} = \begin{pmatrix} 0.836 & 0.056 & 0.024 \\ 0.0352 & 0.024 & 0 \\ 0.0088 & 0 & 0.016 \end{pmatrix}$$

- Eg

$$\begin{aligned} p(Cnt = \mathrm{E}, MT = \mathrm{Eng}) = p(MT = \mathrm{Eng}|Cnt = \mathrm{E})p(Cnt = \mathrm{E}) \\ = 0.95 \times 0.88 \end{aligned}$$

1.2 Probabilistic Reasoning

- The central paradigm of probabilistic reasoning is
  - to identify all variables $x_1, \ldots, x_N$ in the environment
  - make a probabilistic model $p(x_1, \ldots, x_N)$

- Hamburgers
  - Kreuzfeld-Jacob disease (KJ)

$$p(Hamburger\ Eater|KJ) = 0.9$$
$$p(KJ) = 1/100{,}000$$

  - Problem

$$p(KJ|Hamburger\ Eater) = ?$$

- Inspector Clouseau
  - Scene of crime
    victim lies dead, a knife
    suspects: Butler(B), Maid(M)
  - Prior belief of inspector
    Butler is the murderer - 60%
    Maid is the murderer - 20%

- Mathematical formulation

$$\text{dom}(B) = \text{dom}(M) = \{\text{murderer}, \text{not murderer}\}, \text{dom}(K) = \{\text{knife used}, \text{knife not used}\}$$

$$p(B = \text{murderer}) = 0.6, \qquad p(M = \text{murderer}) = 0.2$$

$$\begin{aligned}
p(\text{knife used}|B = \text{not murderer}, \quad & M = \text{not murderer}) &= 0.3 \\
p(\text{knife used}|B = \text{not murderer}, \quad & M = \text{murderer}) &= 0.2 \\
p(\text{knife used}|B = \text{murderer}, \quad & M = \text{not murderer}) &= 0.6 \\
p(\text{knife used}|B = \text{murderer}, \quad & M = \text{murderer}) &= 0.1
\end{aligned}$$

- $p(B|K) = ?$
- $p(B = \text{murderer}|K = \text{knife used})$

- XOR Gate

A	B	A xor B
0	0	0
0	1	1
1	0	1
1	1	0

- Soft XOR Gate

A	B	$p(C=1 \mid A,B)$
0	0	0.1
0	1	0.99
1	0	0.8
1	1	0.25

$$p(A=1 \mid C=0) = ?$$

1.3 Prior, Likelihood and Posterior

- Our interest

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int_\theta p(\mathcal{D}|\theta)p(\theta)}$$

  - Prior - $p(\theta)$
  - Posterior - $p(\theta|\mathcal{D})$
  - Likelihood - $p(\mathcal{D}|\theta)$