Machine Learning for Engineering and Science Applications

Independence

Conditional Independence

Chain rule of probability

Independence

- **Independent random variables** – Two random variables ***X*** and ***Y*** are said to be *statistically independent* if and only if

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

- More precisely, X and *Y* are independent iff

$$p(X = x, Y = y) = p(X = x)p(Y = y) \quad \forall x \in X, y \in Y$$

  - Examples
    - Independent – X: Throw of a dice, Y: Toss of a coin
    - Not independent – X: Height, Y: Weight
- Independence is equivalent to saying

$$p(y|x) = p(y) \text{ OR } p(x|y) = p(x)$$

- Can be seen from product rule $p(x,y) = p(y|x)p(x) = p(x)p(y)$

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

- Independence is denoted by $x \perp y$

Conditional Independence

- Two random variables ***X*** and ***Y*** are said to be *independent given z* if and only if

$$\boxed{p(x, y \mid z) = p(x|z)p(y|z)}$$

- More precisely, *X* and *Y* are independent given Z iff

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z)p(Y = y|Z = z) \quad \forall x \in X, y \in Y, z \in Z$$

  - Examples
    - *Ind & Cond Ind* – **X**: Throw of a dice, **Y**: Toss of a coin, **Z** : Card from deck
    - *Not Ind* BUT *Cond Ind* – **X**: Height, **Y**: Vocabulary, **Z** : Age
    - *Ind* BUT *Cond Not Ind* – **X**: Dice Throw 1, **Y**: Dice Throw 2, **Z** : Sum of dice

Conditional Probability – Chain rule

- Recall Product Rule : $p(x, y) = p(x|y)p(y)$

Consider $p(x, y, z) = p(x, a)$ where a is the event (y, z)

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

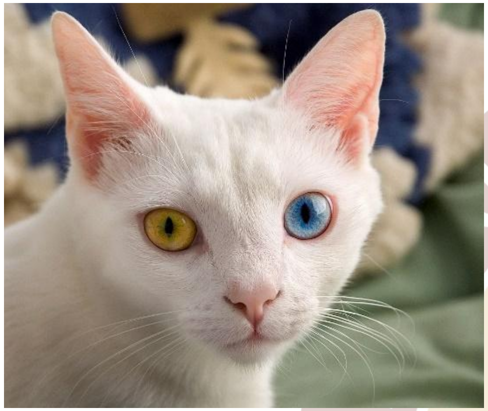
- In general,
- $P\left(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right) = P\left(x^{(1)}\right)P\left(x^{(2)}\middle|x^{(1)}\right) \ldots P\left(x^{(n)}\middle|x^{(1)}, \ldots, x^{(n-1)}\right)$

i.e.

$$\boxed{P\left(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right) = P\left(x^{(1)}\right)\Pi_{i=2}^{n} P\left(x^{(i)}\middle|x^{(1)}, \ldots, x^{(i-1)}\right)}$$

Chain rule of (conditional) probability

One context for conditional probabilities



- Images may be thought of as a collection of pixels $x^{(1)}, x^{(2)}, \dots, x^{(n)}$
- The probability of a particular image may be thought of as joint probability

$$P\left(x^{(1)}, x^{(2)}, \dots, x^{(n)}\right)$$

- Chain rule along with conditional independence can be used to estimate probabilities of the occurrence of images

Keith Kissel - http://flickr.com/photos/74419347@N00/107382056
https://en.wikipedia.org/wiki/Odd-eyed_cat#/media/File:June_odd-eyed-cat_cropped.jpg