Conditional Probability

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

"A occuring given B has already occured"

example:

e(B) - raining $P(\underset{u}{A}|\underset{r}{B})$

e(A) - umbrella

P taking umbella given its raining

Joint Probability

$P(A\cap B)$

For Independent events

$P(A\cap B) = P(A)\times P(B)$

For dependent events

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

"probabiity of 2 events occuring together"

example: getting head on coin & rolling 4 in die

$P(h\cap r_4) = P(h)\times P(r_4)$

$= \left(\frac{1}{2}\right)\times\left(\frac{1}{6}\right) = 1/12$

Marginal Probability

discrete

$$P(A) = \sum_b P(A\cap B=b)$$

continous

$$P(A) = \int_{-\infty}^{\infty} P(A\cap B=b)\,db$$

"Probability of a single event regardless of other events"

example e A = raining

e B = temp > 20°

¬B = temp < 20°

$P(A) = P(A\cap B) + P(A\cap \neg B)$

Confusion Matrix:

Predicted Class

Actual Class		Positive	Negative	
	Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Sensitivity ⇒ ability to identify all positive cases

Specificity ⇒ ability to identify all negative cases

Accuracy ⇒ overall correctness of classification Model

Precision ↳ accuracy of +ve predictions

$$\boxed{\text{Surprise (Information)} \propto \frac{1}{\text{probability}}}$$

$p \uparrow$, surprise $\downarrow$

$p \downarrow$, surprise $\uparrow$

¢ — 1: $p(\text{heads}) = \dfrac{1}{p(\text{heads})} = \dfrac{1}{1} = 1$ but there is no surprise

¢ — 0: $p(\text{tails}) = \dfrac{1}{p(\text{tails})} = \dfrac{1}{0}$ undefined

$\therefore$ Introducing $\log_2$ → when 2 outputs to bring linearity relation

¢ — 1: $p(\text{heads}) = \log\left(\dfrac{1}{p(\text{heads})}\right) = \log(1) = 0$

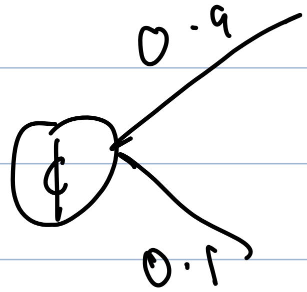
¢ — 0: $p(\text{tail}) = \log\left(\dfrac{1}{p(\text{tail})}\right) = \log\left(\dfrac{1}{0}\right)$

$= \log 1 - \log 0$

$=$ undefined

OK as surprise of something that never happens

example:



$p(\text{head}) = \log_2\left(\frac{1}{0.9}\right) = \log_2\left(\frac{10}{9}\right)$

$= 0.15$

$p(\text{tail}) = \log_2\left(\frac{1}{0.1}\right) = \log_2(10)$

$= 3.32$

Flipping coin 3 times

H, H, T

0.9, 0.9, 0.1

$$= \log_2\left(\frac{1}{0.9 \times 0.9 \times 0.1}\right) \overset{\log 1}{=} -\log_2(0.9 \times 0.9 \times 0.1)$$

$$= -\left[\log_2(0.9) + \log_2(0.9) + \log_2(0.1)\right]$$

$$= -\left[0.15 + 0.15 + 3.32\right]$$

	head	tail
$p(x)$	0.9	0.1
$\log_2\left(\frac{1}{p_x}\right)$	0.15	3.32

Total surprise after flipping coin 100 times

$$= \frac{(0.9 \times 100) \times 0.15 + (0.1 \times 100) \times 3.32}{}$$

surprise for 100 coin flips $= 46.7$

$\boxed{\div \text{ by No of coin tosses}}$

$$= \frac{(0.9 \times 1\cancel{00}) \times 0.15 + (0.1 \times 1\cancel{00}) \times 3.32}{\text{No of coin tosses} = 1\cancel{00}}$$

E (surprise)

On average we expect the surprise to be 0.467 every time we flip coin $= 0.467$

→ entropy (expected surprise every time coin is flipped)

$$= \underset{p(x_i)}{(0.9} \times \underset{\log_2 1/p(x)}{0.15)} + \underset{p(x_i)}{(0.1} \times \underset{\log_2\left(\frac{1}{p_{x_i}}\right)}{3.32)}$$

$$\boxed{H(X) = \sum p(x) \cdot I(x)}$$

$$H(X) = \sum p(x) \log_2 \frac{1}{p(x)} \quad = \sum p_x (I)$$

$$\boxed{\text{Entropy} = -\sum p(x) \log_2 p(x)} \quad [\text{bits}]$$

$$\boxed{\text{Joint Entropy} = H(x,y) = -\sum p(x,y)\log_2(p(x,y)) \text{ bits}}$$

Conditional Entropy =

$$\boxed{H(y|x) = -\sum_{\forall y} p(y|x)\log_2(p(y|x)) \text{ bits}}$$