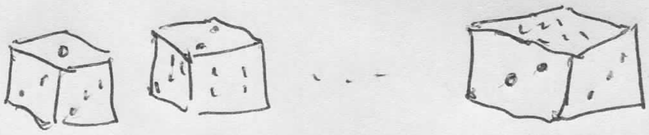
Probability

Sample space: all possible worlds
- mutually exclusive
- exhaustive

} called Ω

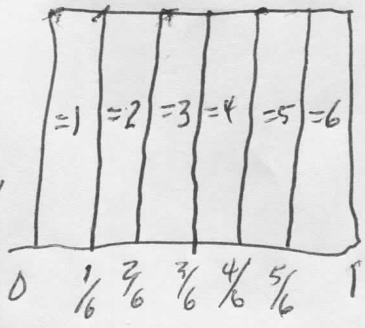
e.g., the rolls of a die



ω_1 ω_2 ... ω_6

why rectangular?

Ω



areas are equal means prob. = 1/6

Probability model: associates a probability to each world

e.g., $p(\omega_i) = \frac{1}{6}$

$$0 \leq \text{Prob}(\omega) \leq 1 \quad \forall \omega \in \Omega$$

$$\sum_{\omega \in \Omega} \text{Prob}(\omega) = 1$$

events: characterized by a set of worlds which make event true

characterized as a proposition ϕ

e.g., consider 2 dice

and $\phi \equiv$ sum of dice is 7

then $Prob(\phi) = \sum_{w_i \text{ makes } \phi \text{ true}} Prob(w_i) = \sum [P(1,6) + P(2,5) + P(3,4) + P(4,3) + P(5,2) + P(6,1)] = \frac{1}{6}$

worlds where ϕ is true

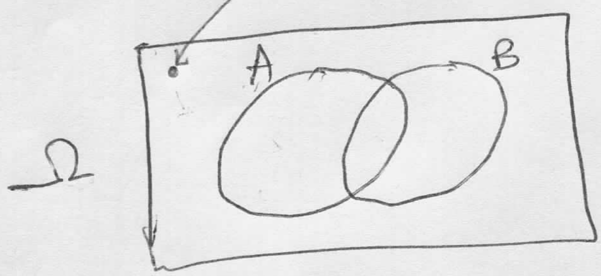
Consider as a sample space diagram

E.g., 2 variables: $A \equiv$ I stay home

$B \equiv$ It's raining

How to draw?

what is a sample?



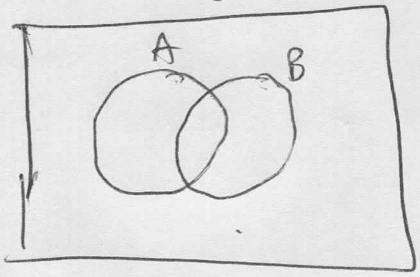
Usually where $P(B) \propto area(B)$

$P(A) \propto area(A)$

what is a world? Both A and B are a function of time!

Usual diagram

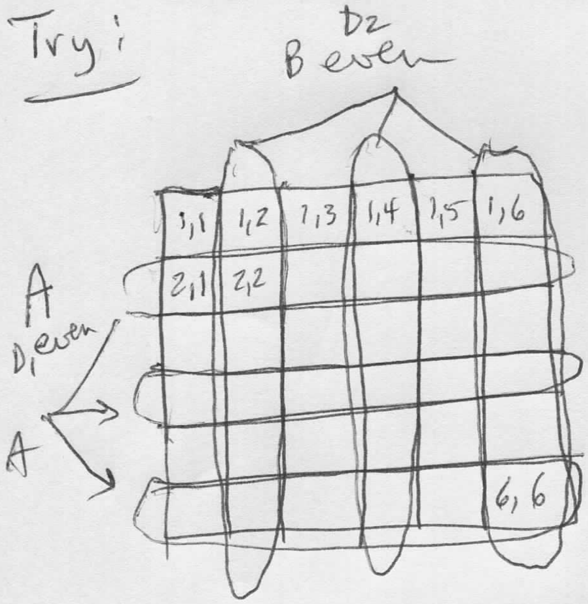
13|3

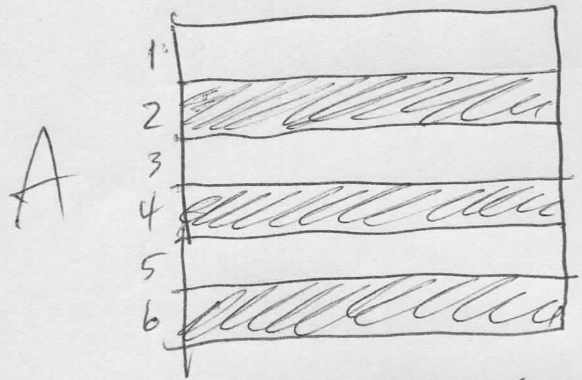


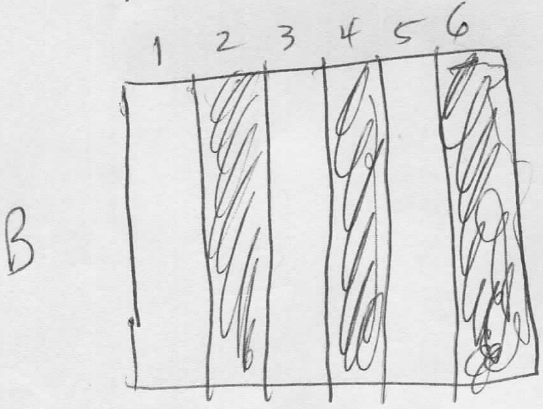
Consider all logical combinations

A	B
0	0
0	1
1	0
1	1

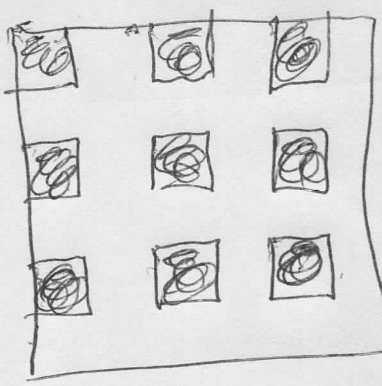
Try:



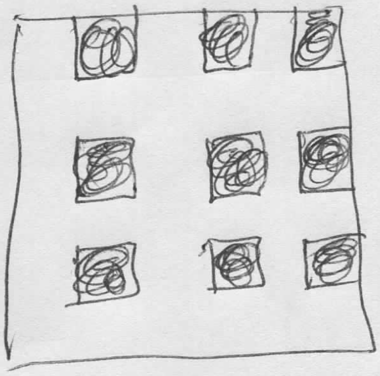


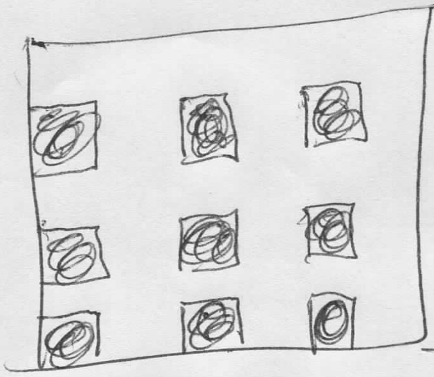


¬A & ¬B

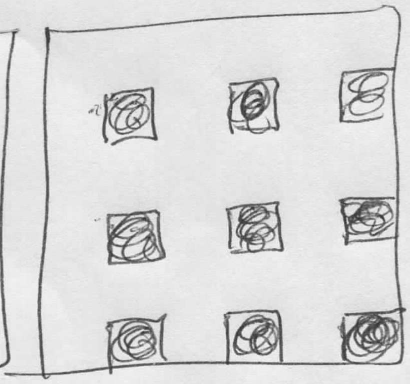


¬A & B





A & ¬B



A & B

Turns out:

$P(\neg A \,\&\, \neg B)$
$P(\neg A \,\&\, B)$
$P(A \,\&\, \neg B)$
$P(A \,\&\, B)$

these probabilities provide the <u>full joint probability distribution</u> .... and

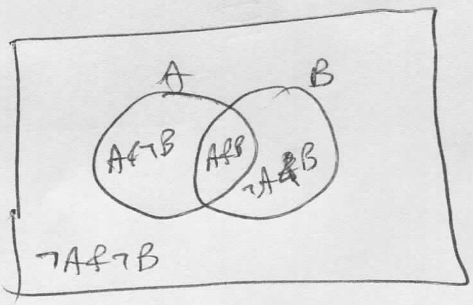
any event's probability can be computed from these.

E.g., $P(A \text{ even}) = P(A \,\&\, \neg B) + P(A \,\&\, B)$

<u>if</u> $P(\neg A \,\&\, \neg B) = P(\neg A \,\&\, B) = P(A \,\&\, \neg B) = P(A \,\&\, B) = 1/4$

<u>then</u> $P(A \text{ even}) = \frac{1}{4} + 1/4 = 1/2$

May be easier to think about as:

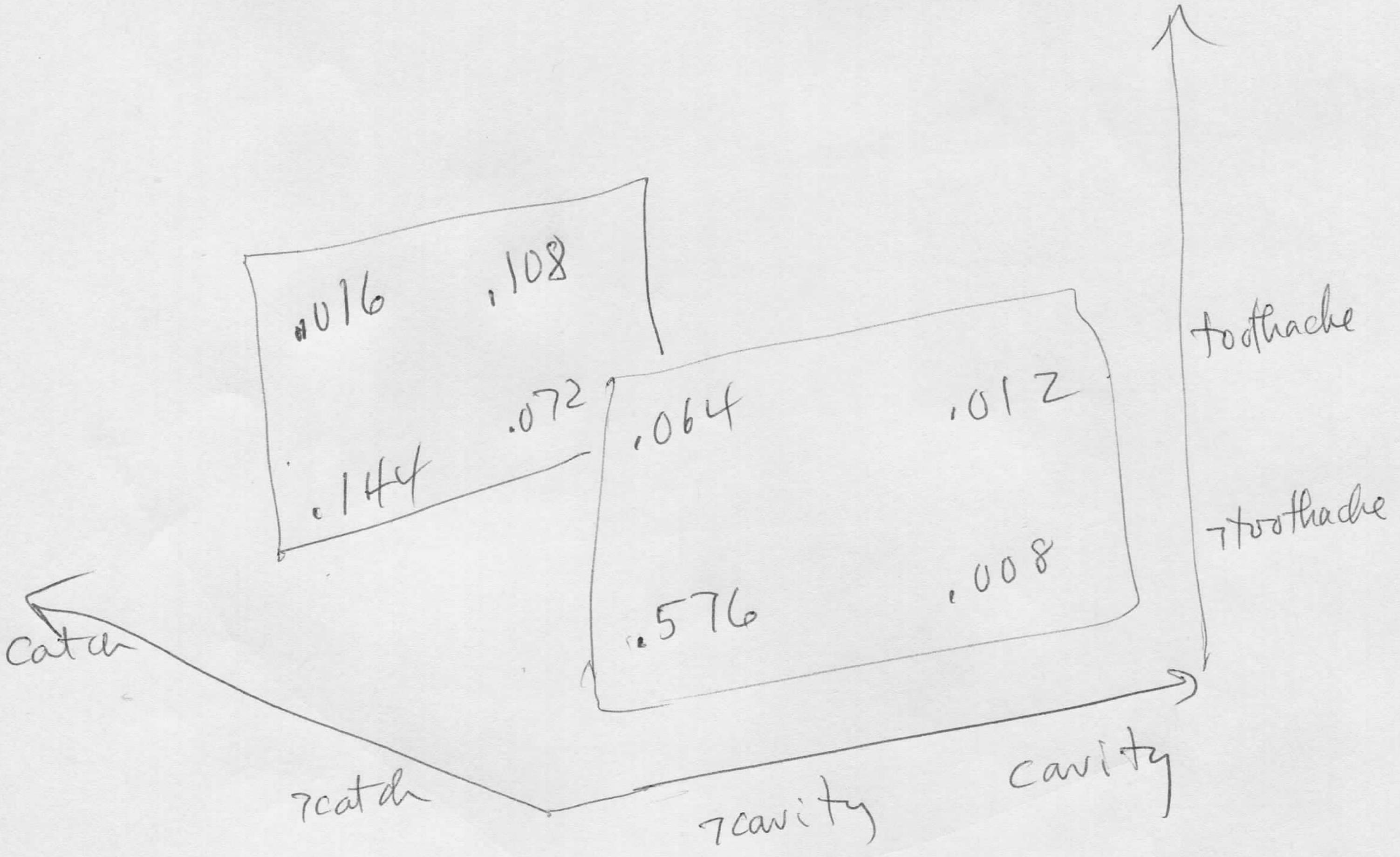


e.g., $P(A \vee B)$

$$= \underbrace{P(A \,\&\, \neg B) + P(\neg A \,\&\, B) + P(A \,\&\, B)}_{\text{all inside the 2 circles}}$$

also called <u>complete conjunction set</u>

Consider a full joint probability distribution for 3 variables



e.g., $P(\neg cavity \,\&\, \neg catch \,\&\, \neg toothache) = .576$

$P(cavity) = .008 + .012 + .072 + .108 = 0.2$

called <u>marginalization</u> or <u>marginal probability</u>

add up over all complete conjunctions that make proposition true

Then can find:

$$P(A) = P(A \,\&\, \neg B) + P(A \,\&\, B)$$
$$P(B) = P(\neg \,\&\, B) + P(A \,\&\, B)$$

} know these

$$P(A \vee B) = P(A) + P(B) - P(A \,\&\, B)$$

why

because $P(A \,\&\, \neg B) + \underline{P(A \,\&\, B)} + P(\neg A \,\&\, B) + \underline{P(A \,\&\, B)}$

but $P(A \vee B) = P(A \,\&\, \neg B) + P(\neg A \,\&\, B) + P(A \,\&\, B)$

& since there's one $P(A \,\&\, B)$ too many subtract one

Probabilities so far are <u>unconditional</u> or <u>prior</u>

But we want ~~to~~ use visual features to classify!

How to use evidence?

$P(\text{class} \mid \text{features}) \equiv$ prob. of class given features

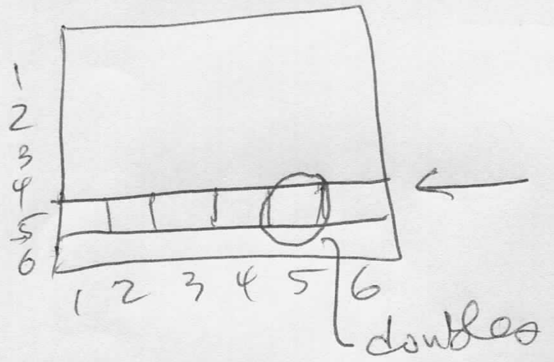
<u>Defined as</u>:

$$P(A \mid B) = \frac{P(A \,\&\, B)}{P(B)}$$

E.g.,

$$P(\text{doubles} \mid Die_1 = 5) = \frac{P(\text{doubles} \wedge Die_1 = 5)}{P(Die_1 = 5)}$$

$$= \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$



$\leftarrow$ $Die_1 = 5$ Given this, only 1 in 6 chance of doubles

---

Can re-write as

$$P(A \mid B)\,P(B) = P(A \,\&\, B)$$
$$P(B \mid A)\,P(A) = P(A \,\&\, B)$$

} product rule

Also

$$\Rightarrow \quad P(B \mid A) = \frac{P(A \mid B)\,P(B)}{P(A)}$$

Bayes' rule

For us

$$P(\text{class} \mid \text{features}) = \frac{P(\text{features} \mid \text{class})\,P(\text{class})}{P(\text{features})}$$

Consider Euler number

A : 16/17 have Euler # = 0

$$P(A \mid Euler == 0) = \frac{P(Euler == 0 \mid A)\, P(A)}{P(Euler == 0)}$$

$$= \frac{\frac{16}{17} \cdot 0.0227}{?}$$ → How can we find this?

e : 67/67 have Euler # 0

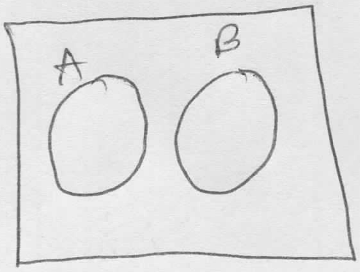
$$P(e \mid Euler == 0) = \frac{P(Euler == 0 \mid e)\, P(e)}{P(Euler == 0)}$$

$$= \frac{1 \cdot 0.0895}{P(Euler == 0)}$$

So, given only Euler #, then
e is more probable

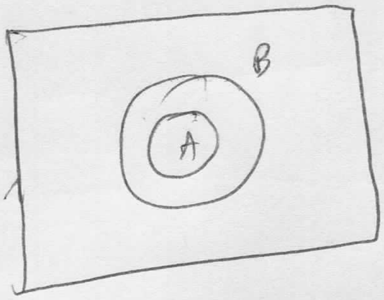
Given A + B, what does it mean for them to be independent?

Case 1



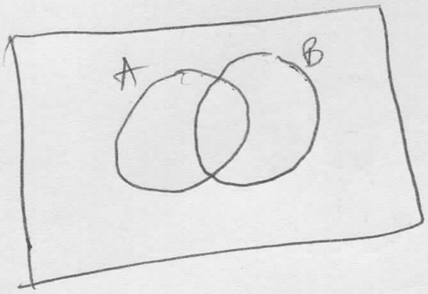
? $P(A) = \frac{1}{4}$
$P(B) = 1/4$

Case 2



? $P(B) = 1/4$
$P(A) = 1/8$

Case 3



? $P(A) = 1/4$
$P(B) = 1/4$
$P(A \& B) = 1/16$

Rule is A + B are <u>independent</u> (pairwise) if:

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

$$\Rightarrow P(A \& B) = P(A|B)\,P(B) = P(A)\,P(B)$$

<u>Case 1</u> ? if B is true, then $P(A) = 0$. X

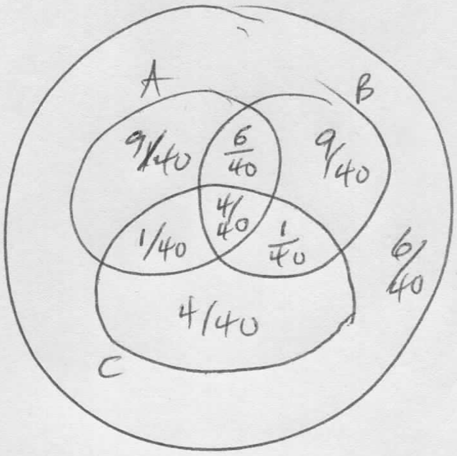
<u>Case 2</u> ? if B is true, then $P(A|B) = \frac{P(A \& B)}{P(B)} = \frac{P(A)}{P(B)}$ X

<u>Case 3</u> ? <u>if</u> B is true, then $P(A|B) = \frac{P(A \& B)}{P(B)} = \frac{1/16}{1/4} = \frac{1}{4}$
$= P(A)$

<u>Observation</u> <u>Mutual independence</u> is when all combinations of n variables satisfy:

$$P(A_1 \,\&\, A_2 \,\&\, \dots \,\&\, A_k) = P(A_1)P(A_2)\dots P(A_k)$$

<u>pairwise independence</u> does <u>NOT</u> imply mutual independence



$P(A) = P(B) = \frac{1}{2}$ $P(C) = \frac{1}{4}$

$P(A|B) = P(A|C) = \frac{1}{2} = P(A)$

$P(B|A) = P(B|C) = 1/2 = P(B)$

$P(C|A) = P(C|B) = 1/4 = P(C)$

$$P(A|BC) = \frac{\frac{4}{40}}{\frac{5}{40}} = \frac{4}{5} \neq \frac{1}{2}$$

$$= \frac{P(A \,\&\, B \,\&\, C)}{P(B \,\&\, C)}$$

Back to Bayes classification

let $\bar{x} = (x_1, x_2, \dots x_n)$ be feature vector

then assigns probabilities to m classes:

$$P(C_k \mid \underbrace{x_1, x_2, \dots x_n}_{x_1 \,\&\, x_2 \,\&\, \dots \,\&\, x_n})$$

Then:

$$P(C_k \mid \bar{x}) = \frac{P(\bar{x} \mid C_k)\,P(C_k)}{P(\bar{x})} \qquad k = 1:m$$

$$P(C_k, x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n, C_k)$$

$$= P(x_1 \mid x_2, x_3, \dots, x_n, C_k)\,P(x_2, x_3, \dots, x_n, C_k)$$

$$= P(x_1 \mid x_2, x_3, \dots x_n, C_k)\,P(x_2 \mid x_3, x_4, \dots, x_n, C_k)\,P(x_3, \dots, x_n, C_k)$$

$$\vdots$$

$$= P(x_1 \mid x_2, x_3, \dots, x_n, C_k)\,P(x_2 \mid x_3, x_4, \dots, x_n, C_k) \dots P(x_{n-1} \mid x_n, C_k)\,P(x_n \mid C_k)\,P(C_k)$$

assume x_i are mutually independent (e.g. Enlai #, height)

then $P(x_i \mid x_{i+1}, \dots, x_n, C_k) = P(x_i \mid C_k)$

and $P(C_k \mid x_1, x_2, \dots, x_n, C_k) \propto P(C_k, x_1, \dots x_n)$

$$= P(C_k)\,P(x_1 \mid C_k) \dots P(x_n \mid C_k)$$

$$= P(C_k) \prod_{i=1}^{n} P(x_i \mid C_k)$$