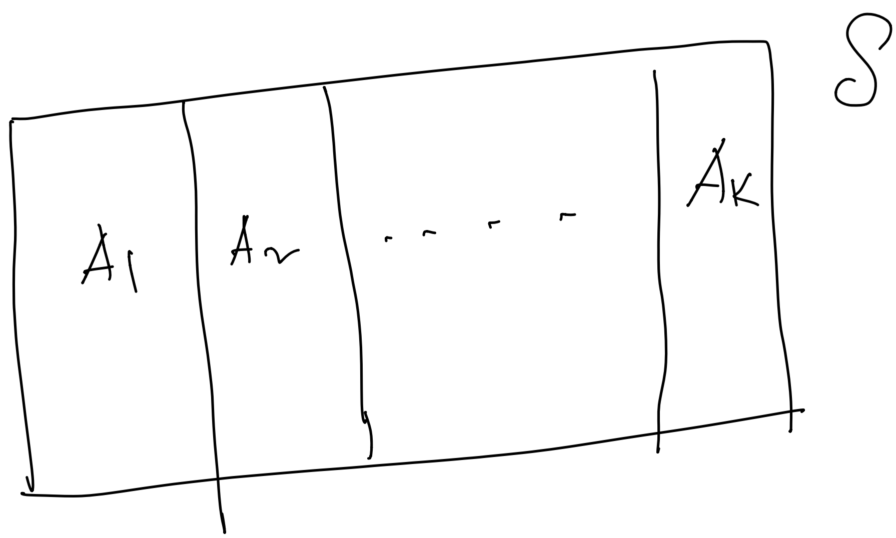
Partitioning of the sample space:

Events $A_1, A_2, \ldots, A_K$ are said to partition the sample space S if:

1. $A_i \cap A_j = \emptyset$ for $i \neq j$

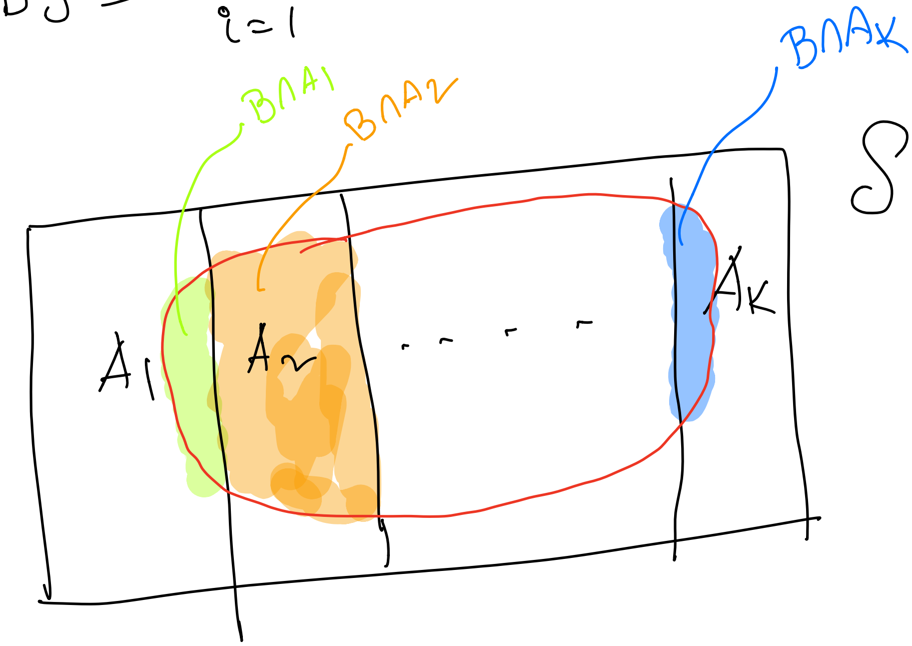
2. $\bigcup_{i=1}^{K} A_i = S$



Law of total probability:

Suppose $A_1, \dots A_K$ forms a partition of the sample space. Then for any event B in the sample space, we have

$$P[B] = \sum_{i=1}^{K} P[B|A_i] P[A_i]$$



Bayes Rule:

Let $\{A_1, A_2, \dots A_K\}$ be a partition of the sample space S, and suppose each of the events $A_1, A_2, \dots A_K$ has nonzero probability. Let B be any event for which $P(B) > 0$. Then for each integer n $(1 \le n \le K)$, we have Bayes formula:

$$P(A_n | B) = \frac{P(A_n) P(B | A_n)}{\sum_{j=1}^{K} P(A_j) P(B | A_j)}$$

Chain Rule:

For any events A and B, we have

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

more generally, for any events $A_1, \dots A_n$,

$$P\left(\cap_{i=1}^{n} A_i\right) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1, A_2) \cdots P\left(A_n \middle| \cap_{i=1}^{n-1} A_i\right)$$

To prove the above, we will use induction on n. The base case is $n=1$. for the base case,

$$P(A_1) = P(A_1)$$

which is trivially true. for the inductive step, let $n>1$ and assume (the inductive hypothesis) that

$$P\left(\cap_{i=1}^{n-1} A_i\right) = P(A_1)\,P(A_2|A_1) \cdots P\left(A_{n-1} \middle| \cap_{i=1}^{n-2} A_i\right)$$

Then,

$$P\left(\cap_{i=1}^{n} A_i\right) = P\left(A_n \cap \left\{ \cap_{i=1}^{n-1} A_i \right\}\right)$$

By the definition of conditional probability,

$$P\left(\cap_{i=1}^{n} A_i\right) = P\left(A_n \mid \cap_{i=1}^{n-1} A_i\right) P\left(\cap_{i=1}^{n-1} A_i\right)$$

Now, by the induction hypothesis,

$$P\left(\cap_{i=1}^{n} A_i\right) = P\left(A_n \mid \cap_{i=1}^{n-1} A_i\right) P(A_1) P(A_2 \mid A_1) \cdots P\left(A_{n-1} \mid \cap_{i=1}^{n-2} A_i\right)$$

This completes the proof by induction.

Practice Problems on probability basics:

1/ Let's define the following events:

Z_D: A randomly selected chip is defective

Z_A: A randomly selected chip was manufactured by A.

Z_B: A randomly selected chip was manufactured by B.

Z_C: A randomly selected chip was manufactured by C.

Now, we want to compute

$P(Z_A|Z_D)$, $P(Z_C|Z_D)$

from the problem statement, we know

$P(Z_D|Z_A) = 0.005$, $P(Z_D|Z_B) = 0.001$

$P(Z_D|Z_C) = 0.01$.

By chain rule of probability,

$$P(Z_A|Z_D) = \frac{P(Z_D|Z_A)\,P(Z_A)}{P(Z_D)}$$

Since Z_A, Z_B and Z_C forms a partition of the sample space, so by law of total probability

$$P(Z_D) = P(Z_D|Z_A)P(Z_A) + P(Z_D|Z_B)P(Z_B) + P(Z_D)Z_C)\,P(Z_C)$$

$$= 0.005 \times 0.5 + 0.001 \times 0.1 + 0.01 \times 0.4$$

$$= 0.0025 + 0.0001 + 0.004$$

$$P(Z_D) = 0.0066$$

Hence,

$$P(Z_A | Z_D) = \frac{0.005 \times 0.5}{0.0066} = 0.3788$$

$$P(Z_C | Z_D) = \frac{0.01 \times 0.4}{0.0066} = 0.606$$

2

Suppose we define the following events:

T: A two headed coin is flipped

F: A fair coin is flipped

B: A biased coin is flipped

H: The flipped coin shows a head

a) Since T, F, and B forms a partition of the sample space, so by law of total probability

$$P(H) = P(H|T)P(T) + P(H|F)\cdot P(F) + P(H|B)P(B)$$

$$= P(T) + \frac{1}{2}P(F) + pP(B)$$

$$= \frac{1}{3} + \frac{1}{6} + \frac{1}{3}p$$

$$\therefore \quad P(H) = \frac{1}{2} + \frac{1}{3}p$$

Since $0 \le p \le 1$, so

$$\max P(H) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

b) By the definition of conditional probability,

$$P(F \mid H) = \frac{P(H \mid F) \cdot P(F)}{P(H)}$$

$$P(F \mid H) = \frac{1/2 \cdot 1/3}{1/2 + 1/3\,p} = \frac{1/6}{1/2 + 1/3\,p}$$

Since $0 \leq p \leq 1$, so $P(F|H)$ is maximized when $p = 0$

$\max P(F|H) = 1/3$

$P(F|H)$ is minimized when $p = 1$

$\min P(F|H) = \frac{1/6}{5/6} = 1/5$

3

suppose we define the following events

E_1: Person selected has tuberculosis

E: Person's chest X-Ray is diagonsed as positive i.e as showing tuberculosis

Since E_1 and E_1^C forms a partition of the sample space, so by Bayes rule we have

$$P(E_1|E) = \frac{P(E_1)\,P(E|E_1)}{P(E|E_1)\cdot P(E_1) + P(E|E_1^c)\cdot P(E_1^c)}$$

$$= \frac{(0.001)(0.9)}{(0.001)(0.9) + (0.01)(0.999)}$$

$$= 0.083$$

$$\therefore P(E_1|E) = 0.083$$

Although the X-Ray test is fairly reliable, we have found that only slightly more than 8% of those with positive X Rays turn out to have tuberculosis. The results of such calculations must be taken into account when large-scale medical diagnostic tests are planned.

Practice problem on linear algebra basics

(a) Let λ_i be an eigenvalue of A with corresponding eigenvector v_i. Then,

$$Av_i = \lambda_i v_i$$

multiplying both sides on the left by v_i^T. we get,

$$v_i^T A v_i = \lambda_i v_i^T v_i$$

$$\Rightarrow v_i^T A v_i = \lambda_i \|v_i\|_2^2$$

Since $b^T A b > 0$ for all $b \in \mathbb{R}^n$, so

$$v_i^T A v_i = \lambda_i \|v_i\|_2^2 > 0$$

Since $\|v_i\|_2^2 > 0$, so $\lambda_i > 0$.

$\therefore$ All eigenvalues of A are positive.

(b) Let λ_i be an eigenvalue of A with corresponding eigenvector v_i. Then,

$$A v_i = \lambda_i v_i$$

multiplying both sides to the left by A^T, we get

$$A^T A v_i = \lambda_i A^T v_i$$

Since A is an orthogonal matrix, so

$$A^T A = A A^T = I$$

Hence,

$$v_i = \lambda_i A^T v_i$$

Taking the 2-norm of both sides

$$\|v_i\|_2^2 = |\lambda_i|^2 \|A^T v_i\|_2^2$$

Now,

$$\|A^T v_i\|_2^2 = (A^T v_i)^T (A^T v_i)$$

$$= v_i^T A A^T v_i = \|v_i\|_2^2$$

Hence, we have

$$\|v_i\|_2^2 = |\lambda_i|^2 \|v_i\|_2^2$$

$\therefore$ $|\lambda_i| = 1$. Hence, all eigenvalues of A have norm 1.

(c) $A \in \mathbb{R}^{m \times n}$ has the following SVD

$$A = U \Sigma V^T$$

Now,

$$AA^T = (U \Sigma V^T)(U \Sigma V^T)^T$$

$$= U \Sigma V^T V \Sigma^T U^T$$

$$AA^T = U \Sigma \Sigma^T U^T$$

Now, $\Sigma = \begin{bmatrix} \sigma_1 & \sigma_2 & \cdots & \sigma_r & 0 & \cdots & 0 \end{bmatrix}$

So, $\Sigma \Sigma^T = \begin{bmatrix} \sigma_1^2 & & & & & \\ & \sigma_2^2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r^2 & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$

since $\Sigma \Sigma^T$ is a diagonal matrix and U is an orthogonal matrix so the eigendecomposition of the symmetric matrix AA^T is

$$U \Sigma \Sigma^T U^T$$

Hence,

$$\lambda_i(AA^T) = \sigma_i^2(A)$$

2

a) $(A^T)^T A^T$

$= AA^T$

$= I$ (Since A is orthogonal)

Also,

$A^T (A^T)^T$

$= A^T A$

$= I$ (Since A is orthogonal)

Hence, A^T is also orthogonal

b) $(AB)^T(AB)$

$= B^T A^T A B$

$= B^T I B$ (since A is orthogonal)

$= B^T B$

$= I$ (since B is orthogonal)

Also, $AB(AB)^T$

$= A B B^T A^T$

$= A I A^T$ (Since B is orthogonal)

$= A A^T$

$= I$ (since A is orthogonal)

$\therefore$ AB is orthogonal

c) Let $A = B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Then A and B are both orthogonal,

but

$$A+B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

is not orthogonal.

d) Suppose the column vectors of A are orthonormal. Hence

$$a_i^T a_i = 1 \quad \forall i$$

$$a_i^T a_j = 0 \quad i \neq j$$

which implies

$$A^T A = I$$

Since $A^{-1} = A^T$, so A is an orthogonal matrix. From (a), we know that A^T is also orthogonal. Since A^T is also orthogonal, so

$$(A^T)^T A^T = I$$

The above relation implies A^T has orthonormal columns, meaning that A has orthormal rows.