Mathematical and statistical Literacy for Machine Learning

Mathematics – in a single page

Mathematics deal with Numbers.
Numbers are combined using Operators,
To form Arithmetic.
Algebra is the generalization of Arithmetic.
An Algebraic Equation with Power of one
Ends up as Linear Equation,
When Power becomes two, It is Quadratic,
When Power becomes three, It is Cubic,
All else are Polynomials.
An Algebraic equation can be morphed
Into Geometry,thanks to Descartes
A coefficient in a Algebraic equation,
when it becomes a constant,
becomes Ordinary Algebraic Equation.
When it becomes Derivative, It becomes
a Differential Equation
When it becomes Partial Derivative, It
becomes a Partial Differential equation
When the coefficients are Probabilistic,
It becomes Stochastic Differential Equation
If you do not solve algebraic equations,
in its closed form, resort to numerical methods.
Thus,the Whole world can be reduced into
Algebraic Equation.
Long Live Algebra,Long Live Mathematics

A Challenge – What is the basis for this "cocksureness" ?

Sunday, December 01, 2013

A Rs. 1,00,00,000 Offer from me , If you are able to find a triplets (3 #'s satisfying a mathematical property)

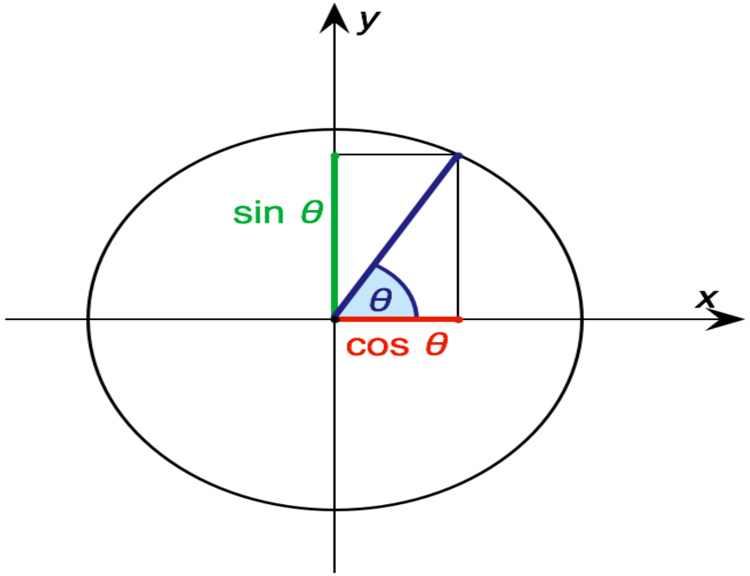
When I learned to write computer programs,finding Pythagorean triplets was one of the first programs written by me. One can easily find lot of triplets which satisfy Pythagorean triangle equation ,
x^2 + y^2 = z^2.

Some examples are , 4^2 + 3^2 = 5^2
12^2 + 5^2 = 13^2

The challenge is to find out triplets for any number n , provided n > 2 which satisfies the equation
x^n + y^n = z^n.

If you are able to find a integer solution, you will get this reward !

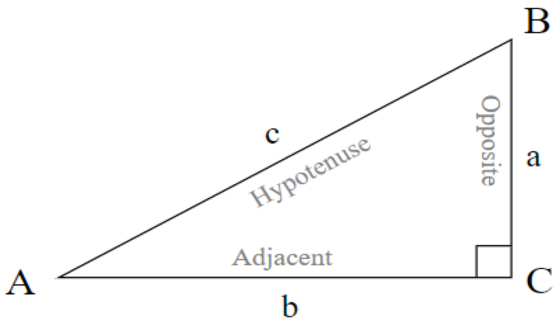
Derivation of Trigonometric Formula



$$\sin A = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{a}{c}.$$

$$\cos A = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{b}{c}.$$

$$\tan A = \frac{\text{opposite}}{\text{adjacent}} = \frac{a}{b} = \frac{a/c}{b/c} = \frac{\sin A}{\cos A}.$$

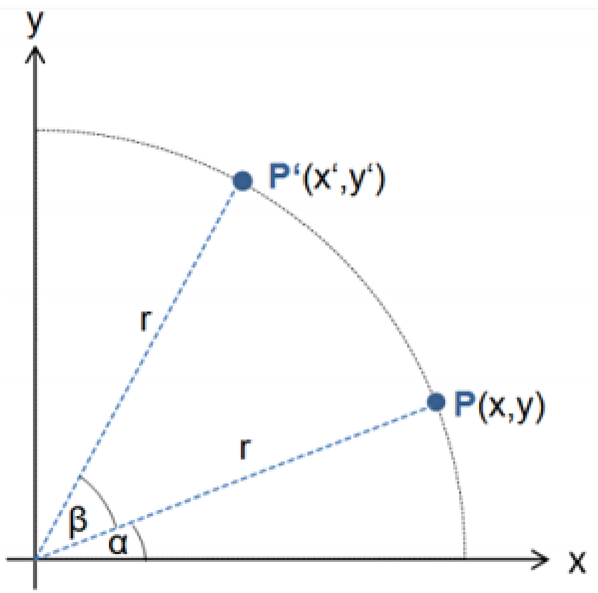


$$\csc A = \frac{1}{\sin A} = \frac{\text{hypotenuse}}{\text{opposite}} = \frac{c}{a},$$

$$\sec A = \frac{1}{\cos A} = \frac{\text{hypotenuse}}{\text{adjacent}} = \frac{c}{b},$$

$$\cot A = \frac{1}{\tan A} = \frac{\text{adjacent}}{\text{opposite}} = \frac{\cos A}{\sin A} = \frac{b}{a}.$$

2D Rotation Scenario



$$x = r * \cos\alpha$$
$$y = r * \sin\alpha$$

$$x' = r * \cos(\alpha + \beta)$$
$$y' = r * \sin(\alpha + \beta)$$

$$\begin{aligned} x' &= r * \cos(\alpha + \beta) \\ &= r * (\cos\alpha * \cos\beta - \sin\alpha * \sin\beta) \\ &= r * \cos\alpha\cos\beta - r * \sin\alpha\sin\beta \\ &= x * \cos\beta - y * \sin\beta \end{aligned}$$

$$\begin{aligned} y' &= r * \sin(\alpha + \beta) \\ &= r * (\sin\alpha * \cos\beta + \cos\alpha * \sin\beta) \\ &= r * \sin\alpha\cos\beta - r * \cos\alpha\sin\beta \\ &= y * \cos\beta + x * \sin\beta \end{aligned}$$

$$[x' \quad y'] = [x \quad y]\begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix}$$

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}.$$

A Mathematical Trivia

Arvind Kejriwal & A "nerdy" joke re-visited

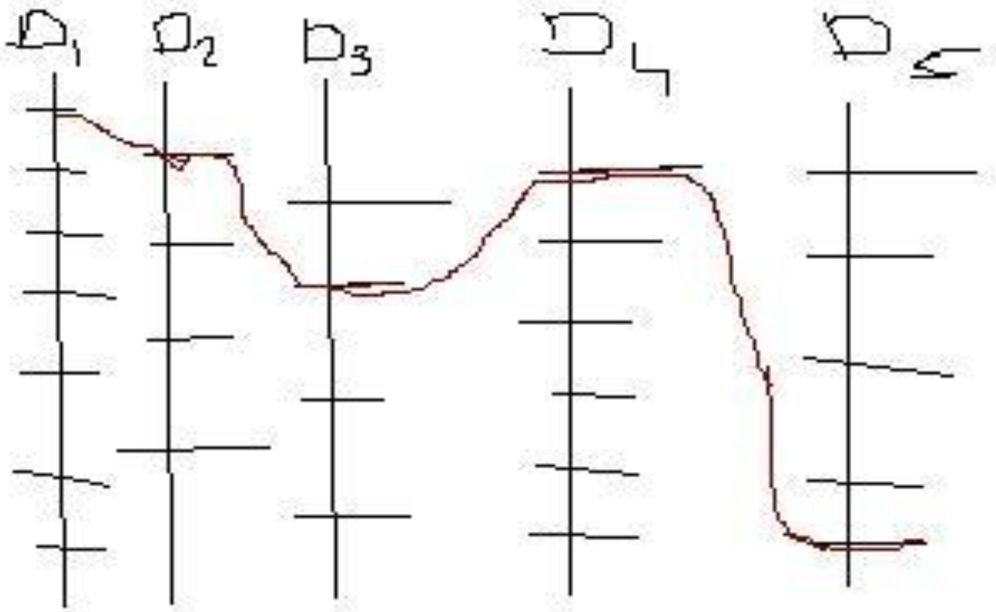
Years ago, from a book, "Secured Programming Cookbook", I encountered the following sentence,
"**there are 10 types of people in this world, those who understand binary and those who dont**". It took some time for me to understand the joke inside. **The number 10, If interpreted as binary, is decimal two**.

Arvind Kerjriwal, in April,2014 told the following

"AAP will Win 100 LS seats, No question of tie-ups"

A joke appeared on the Facebook, to interpret the numbers in binary. 100 - in binary, is 4 decimal. Exactly the number of seats won by AAP in 2014.

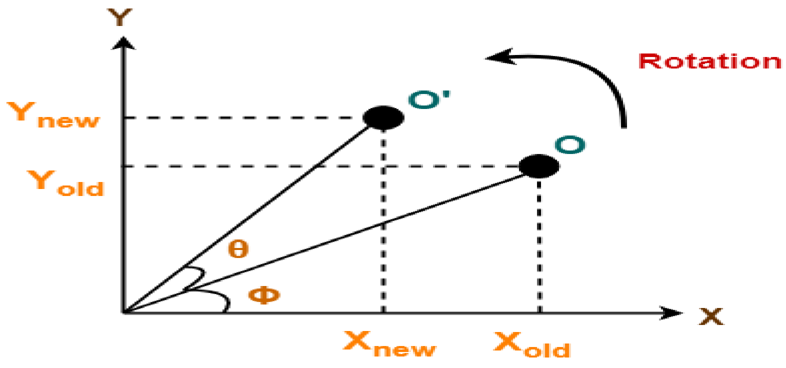
How to Visualize more than three dimensions



A point can represent one dimension (number line) , a line can represent 2 dimension (a plane) and a cube three dimension (space) . when it comes to four and above , cartesian geometric depiction of dimension fails .

2.161 "There must be something identical in a picture and what it depicts , to enable one to be a picture of the other at all"

2D Rotation



2D Rotation in Computer Graphics

$$\begin{bmatrix} X_{new} \\ Y_{new} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \times \begin{bmatrix} X_{old} \\ Y_{old} \end{bmatrix}$$

Rotation Matrix

$$\begin{bmatrix} X_{new} \\ Y_{new} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} X_{old} \\ Y_{old} \\ 1 \end{bmatrix}$$

Rotation Matrix

(Homogeneous Coordinates Representation)

Fermat’s Last Theorom

Words which gave Mathematicians Sleepless nights for Centurites

The Celebrated French Mathematician wrote the following in one of his book.

"It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain."

What essentially, he told was

x^3 <> y^3 + y^3 or x^4 <> y^4 + z^4,
in general, x^n + y^n <> z^n for n > 2

We are familiar with Pythagores Therom, whcih states that,

"In a right triangle, hypotenuse (z) squared is equivalent to base (x) squared added to opposite side(y)"
Algebraically, It can be written as,
z^2 = x^2 + y^2.

Diophantine equations are generalization of Pythagoras theorem where factors (x,y,z) are integral numbers. What Fermat essentially told us was , **For a Diophantine equation with powers greater than 2, there is no solutions.**

Tony Wiles, British mathematician proved it finally, in 1994. Counless mathematicians have attempted a proof and Finally, a solution was found in 1994 (proposed in 1993), 300 years after Fermat's death.

Arithematic Mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\bar{x} = \frac{\sum x}{n}$$

$$\mu = \frac{\sum x}{n}$$

$$\operatorname{E}[X] = \sum_{i=1}^{k} x_i \, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k.$$

$$\operatorname{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

$$\operatorname{E}[X] = \mu.$$

Geometric Mean

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} = \sqrt[n]{a_1 a_2 \cdots a_n}.$$

When $a_1, a_2, \ldots, a_n > 0$

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} = \exp\left[\frac{1}{n}\sum_{i=1}^{n} \ln a_i\right]$$

additionally,

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} = (-1)^m \exp\left[\frac{1}{n}\sum_{i=1}^{n} \ln|a_i|\right]$$

where m is the number of negative numbers.

Harmonic Mean

The harmonic mean H of the positive real numbers $x_1, x_2, \ldots, x_n$ is defined to be

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \left(\frac{\sum_{i=1}^{n} x_i^{-1}}{n} \right)^{-1}.$$

$$1/H(1/x_1 \ldots 1/x_n) = A(x_1 \ldots x_n)$$

Arun Travelled from A to B , at 60 km per hour. While returning, he travelled at 20 km per hour. What is the Average Speed?

Average speed

$$= \frac{\text{Total distance traveled}}{\text{Total time taken}}$$

$$= \frac{2d}{\frac{d}{x} + \frac{d}{y}} = \frac{2d}{\frac{yd + xd}{xy}} = \frac{2dxy}{d(x + y)}$$

$$= \frac{2xy}{x + y} \text{ (harmonic mean of } x \text{ and } y\text{)}$$

Standard Deviation

$$\sigma = \sqrt{\frac{1}{N}\left[(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_N-\mu)^2\right]}, \text{ where } \mu = \frac{1}{N}(x_1 + \cdots + x_N),$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}, \text{ where } \mu = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

The "**Population** Standard Deviation": $$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$$

The "**Sample** Standard Deviation": $$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{x})^2}$$

$$\mathrm{E}[X] = \mu.$$

$$\begin{aligned}
\sigma &= \sqrt{\mathrm{E}[(X-\mu)^2]} \\
&= \sqrt{\mathrm{E}[X^2] + \mathrm{E}[-2\mu X] + \mathrm{E}[\mu^2]} \\
&= \sqrt{\mathrm{E}[X^2] - 2\mu\,\mathrm{E}[X] + \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - 2\mu^2 + \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - \mu^2} \\
&= \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2}
\end{aligned}$$

Mean Absolute Deviation

The mean absolute deviation of a set $\{x_1, x_2, ..., x_n\}$ is

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - m(X)|.$$

Covariance

$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

$$\boxed{\operatorname{cov}(X,Y) = \operatorname{E}[(X - \operatorname{E}[X])\,(Y - \operatorname{E}[Y])]}$$

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \operatorname{E}[(X - \operatorname{E}[X])\,(Y - \operatorname{E}[Y])] \\
&= \operatorname{E}[XY - X\operatorname{E}[Y] - \operatorname{E}[X]Y + \operatorname{E}[X]\operatorname{E}[Y]] \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y] - \operatorname{E}[X]\operatorname{E}[Y] + \operatorname{E}[X]\operatorname{E}[Y] \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y],
\end{aligned}$$

Correlation

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

$$\rho_{X,Y} = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{E}(X^2) - \mathrm{E}(X)^2} \cdot \sqrt{\mathrm{E}(Y^2) - \mathrm{E}(Y)^2}}$$

Covariance Matrix

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^{\mathrm{T}}$$

$$\mathrm{K}_{X_i X_j} = \operatorname{cov}[X_i, X_j] = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])]$$

$$\mathrm{K}_{\mathbf{X}\mathbf{X}} = \begin{bmatrix} \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])] & \cdots & \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_n - \mathrm{E}[X_n])] \\ \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_2 - \mathrm{E}[X_2])] & \cdots & \mathrm{E}[(X_2 - \mathrm{E}[X_2])(X_n - \mathrm{E}[X_n])] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_1 - \mathrm{E}[X_1])] & \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_2 - \mathrm{E}[X_2])] & \cdots & \mathrm{E}[(X_n - \mathrm{E}[X_n])(X_n - \mathrm{E}[X_n])] \end{bmatrix}$$

$$\boxed{\mathrm{K}_{\mathbf{X}\mathbf{X}} = \operatorname{cov}[\mathbf{X}, \mathbf{X}] = \mathrm{E}[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^{\mathrm{T}}] = \mathrm{E}[\mathbf{X}\mathbf{X}^T] - \mu_{\mathbf{X}}\mu_{\mathbf{X}}^T \quad \textbf{(Eq.1)}}$$

BETA (Finance)

$$\beta = \frac{\mathrm{Cov}(r_a, r_b)}{\mathrm{Var}(r_b)},$$

$$\sigma_a = \sqrt{\mathrm{Var}(r_a)},\ \sigma_b = \sqrt{\mathrm{Var}(r_b)},\ \rho_{a,b} = \mathrm{Cov}(r_a, r_b)/\sqrt{\mathrm{Var}(r_a)\mathrm{Var}(r_b)},$$

$$\beta = \rho_{a,b}\frac{\sigma_a}{\sigma_b}$$

Moment

Moment ordinal	Moment		
	Raw	Central	Standardized
1	Mean	0	0
2	–	Variance	1
3	–	–	Skewness
4	–	–	(Non-excess or historical) kurtosis

EMI

$$P = A \cdot \frac{1-(1+r)^{-n}}{r}$$

$$A = P \cdot \frac{r(1+r)^n}{(1+r)^n - 1}$$

IRR

$$\mathrm{NPV} = \sum_{n=0}^{N} \frac{C_n}{(1+r)^n} = 0$$

If an investment may be given by the sequence of cash flows

Year (n)	Cash flow (C_n)
0	-123400
1	36200
2	54800
3	48100

then the IRR r is given by

$$\mathrm{NPV} = -123400 + \frac{36200}{(1+r)^1} + \frac{54800}{(1+r)^2} + \frac{48100}{(1+r)^3} = 0.$$

In this case, the answer is 5.96% (in the calculation, that is, r = .0596).

Quadratic Equation

$$ax^2 + bx + c = 0$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Dot Product Trivia

Why DotProduct is term wise mutliplication?

V = (ax,ay) and U = (bx,by) , their DotProduct is U.V = (ax*bx , ay*by)

V = (ax*I , ay*J) , U = (bx*I,by*J) , I and J be unit vector on X and Y

Term-wise multiplication yields

U.V = (ax*bx*I*I + ax*by*I*J , ay*bx*J*I + ay*by*J*J)

As I*J=0 , I*I = 1 , J*J=1,J*I= 0 , the whole stuff reduces to

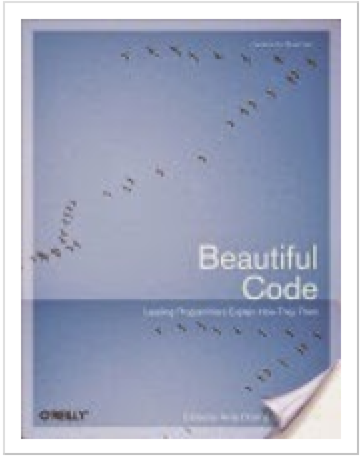
U.V= (ax*bx*1 + 0 , 0+ ay*by*1)

Therefore, U.V = (ax*bx , ay*by)

A Book and its Aftermath

Vaughan Pratt's introduction to Algebra @ SEP

I came across the name Vaughan Pratt in the year 2007/08 from the book, "The Beautiful Code". There is an article by Douglas Crawford on the topic of "Top Down Operator Precedence" Parsing. The code in the article is written using JavaScript, based on a paper by Vaughan Pratt.



The Article on Algebra is written by Vaughan Pratt, from Stanford University. The initial paragraph is beautifully written

"Algebra is a branch of mathematics sibling to geometry, analysis (calculus), number theory, combinatorics, etc. Although algebra has its roots in numerical domains such as the reals and the complex numbers, in its full generality it differs from its siblings in serving no specific mathematical domain. Whereas geometry treats spatial entities, analysis continuous variation, number theory integer arithmetic, and combinatorics discrete structures, algebra is equally applicable to all these and other mathematical domains"

Permutations and Combinations

$$nPk = \frac{n!}{(n-k)!}$$

$$nCk = \frac{n!}{k!(n-k)!} = \frac{nPk}{k!}$$

$$8P2 = \frac{8!}{(8-2)!} = \frac{8!}{6!} = 56$$

$$8C2 = \frac{8!}{2!(8-2)!} = \frac{8P2}{2!} = \frac{56}{2} = 28$$

Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

$$(x+y)^n = \binom{n}{0} x^n y^0 + \binom{n}{1} x^{n-1} y^1 + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n-1} x^1 y^{n-1} + \binom{n}{n} x^0 y^n$$

$$(1+x)^n = \binom{n}{0} x^0 + \binom{n}{1} x^1 + \binom{n}{2} x^2 + \cdots + \binom{n}{n-1} x^{n-1} + \binom{n}{n} x^n,$$

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \prod_{\ell=1}^{k} \frac{n-\ell+1}{\ell} = \prod_{\ell=0}^{k-1} \frac{n-\ell}{k-\ell}$$

Generalized Binomial Theorom

$$
\begin{aligned}
(x+y)^r &= \sum_{k=0}^{\infty} \binom{r}{k} x^{r-k} y^k \\
&= x^r + r x^{r-1} y + \frac{r(r-1)}{2!} x^{r-2} y^2 + \frac{r(r-1)(r-2)}{3!} x^{r-3} y^3 + \cdots .
\end{aligned}
$$

Derivation of E through Binomial Theorom

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n.$$

$$\left(1 + \frac{1}{n}\right)^n = 1 + \binom{n}{1}\frac{1}{n} + \binom{n}{2}\frac{1}{n^2} + \binom{n}{3}\frac{1}{n^3} + \cdots + \binom{n}{n}\frac{1}{n^n}.$$

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}$$

$$\lim_{n\to\infty} \binom{n}{k}\frac{1}{n^k} = \frac{1}{k!}.$$

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots.$$

Logarithm Soup

$$\log_{10}(x) = \frac{\ln(x)}{\ln(10)} \qquad \text{or} \qquad \log_{10}(x) = \frac{\log_2(x)}{\log_2(10)}$$

Rule of 70

$$
\begin{aligned}
(e^r)^p &= 2 \\
e^{rp} &= 2 \\
\ln e^{rp} &= \ln 2 \\
rp &= \ln 2 \\
p &= \frac{\ln 2}{r} \\
\\
p &\approx \frac{0.693147}{r}
\end{aligned}
$$

Arithmetic Progression

$$a_n = a_1 + (n-1)d,$$

$$a_n = a_m + (n-m)d\,.$$

$$S_n = a_1 + (a_1 + d) + (a_1 + 2d) + \cdots + (a_1 + (n-2)d) + (a_1 + (n-1)d)$$

$$S_n = (a_n - (n-1)d) + (a_n - (n-2)d) + \cdots + (a_n - 2d) + (a_n - d) + a_n$$

$$2S_n = n(a_1 + a_n).$$

$$S_n = \frac{n}{2}(a_1 + a_n).$$

$$S_n = \frac{n}{2}[2a_1 + (n-1)d].$$

Geometric Series

$$a,\ ar,\ ar^2,\ ar^3,\ ar^4,\ \ldots$$

$$\sum_{k=1}^{n} ar^{k-1} = ar^0 + ar^1 + ar^2 + ar^3 + \cdots + ar^{n-1}.$$

$$\begin{aligned}(1-r)\sum_{k=1}^{n} ar^{k-1} &= (1-r)(ar^0 + ar^1 + ar^2 + ar^3 + \cdots + ar^{n-1}) \\ &= ar^0 + ar^1 + ar^2 + ar^3 + \cdots + ar^{n-1} - ar^1 - ar^2 - ar^3 - \cdots - ar^{n-1} - ar^n \\ &= a - ar^n\end{aligned}$$

$$\sum_{k=1}^{n} ar^{k-1} = \frac{a(1-r^n)}{1-r}.$$

Derivation of EMI

$P_1 = P \times (1 + r) - E$

$P_2 = P_1 \times (1 + r) - E$

$P_2 = (P \times (1 + r) - E) \times (1 + r) - E$

$P_2 = P \times (1 + r)^2 - E \times ((1 + r) + 1)$

$P_2 = P \times t^2 - E \times (1 + t)$

$P_i = P \times t^i - E \times (1 + t + t^2 + \ldots + t^{i-1})$

$P_n = P \times t^n - E \times (1 + t + t^2 + \ldots + t^{n-1}) = 0$

$P \times t^n = E \times (1 + t + t^2 + \ldots + t^{n-1})$

$P \times t^n = E \times (t^n - 1)/(t - 1)$

$E = P \times t^n \times (t - 1)/(t^n - 1)$

$E = P \times r \times (1 + r)^n/((1 + r)^n - 1)$

$$E = P \cdot r \cdot \frac{(1+r)^n}{((1+r)^n - 1)}$$

For investing 100, if u get 60 rs each for two years, what is IRR

$$100 = \frac{60}{1+r} + \frac{60}{(1+r)^2}.$$

$$60x^2 + 60x - 100 = 0,$$

$$x = \frac{-60 \pm \sqrt{60^2 + 4(60)(100)}}{120}.$$

$$x = \frac{\sqrt{27,600} - 60}{120} \approx .8844.$$

$$1 + r^* \approx \frac{1}{.8844} \approx 1.131.$$