ML Math

1. Linear Algebra & Matrix
2. Statistics
3. Geometry.
4. Calculas
5. Probablity & Distributions
6. Regression
7. Dimensionality Reduction

1. Linear Algebra & Matrix

→ Matrix

$[M] \rightarrow m \times n$

m → no of rows, n → columns

$[M]^T$

→ Determinant of Matrix

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $|A| = ad - bc$

$A = \begin{bmatrix} a_{11} & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ $|A| = a(-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + b(-1)^{1+2} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + c(-1)^{1+3} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$

1. Squar Matrix = $m = n$
2. Symmetric Matrix = $A^T = A$
3. Skew Symmetric Matrix = $A^T = -A$
4. Diagonal Matrix → other than diagonal elements all are zero
5. Identity Matrix → Diagonal elements are 1 rest everything 0.
6. Orthogonal Matrix $AA^T = I$
7. ~~Ide~~ Idempotent → $A^2 = A$
8. Involuntory → $A^2 = I$
9. Null → All elements are zero
10. Uper Triangular
11. Lower Triangular.

→ Adjoint of Matrix → Transpose of the cofactor of Matrix A

→ Inverse of Matrix → $\frac{Adj\,A}{|A|}$

→ Trace of a Matrix → sum of diagonal elements of matrix

Properties of Adjoint

1. $A(Adj\,A) = |A| I_n$
2. $|Adj\,A| = |A|^{n-1}$
3. $Adj\,(kA) = k^{n-1} Adj(A)$
4. $|adj\,(adj(A))| = |A|$ ^(n-1)^2
5. $adj\,(adj(A)) = |A|$ ^(n-2)*A
6. $adj\,(I) = I$

Cummulative Property.

$A + B = B + A$

Associative Property

$A + (B + C) = (A + B) + C$

Distributive Property

$c[A + B] = c[A] + c[B]$

Adding 2 matrices in Python

X = [[1,2,3], [4,5,6], [7,8,9]]
Y = [[9,8,7], [6,5,4], [3,2,1]]

①
```
result = [map(sum, zip(*i))
          for i in zip(X, Y)]
```

②
```
result = np.array(X) + np.array(Y)
```

③ → library for symbolic mathematics.
```
from sympy import Matrix
#create Matrix obj
matrix_x = Matrix(X)
matrix_y = Matrix(Y)
result = matrix_x + matrix_y
```

Multiplication

Note zip(X) → takes as its [1,2,3]...
zip(*X) → takes columns [1,4,7], [2,5,8]...

①
```
result = [[sum(a*b for
           a,b in zip(A_row, B_col))
           for B_col in zip(*B)]
           for A_row in A]
```

②
```
result = np.dot(A, B)
```
↳ define or initialize to 3x3 0 matrix

Note: numpy.add(X, Y)
numpy.subtract(X, Y)
numpy.divide(X, Y)
numpy.multiply(X, Y) → element wise not actual product of matrix

numpy.dot(X, Y) → Actual product of 2 matrices.

np.linalg.inv(X) → inverse of matrix X

Transpose

Z = zip(*X)
or
Z = numpy.transpose(X)
or
Z = X.T

Determinant

det = np.linalg.det(X)
round(det)

Normal of a matrix =
$\sqrt{1^2+2^2+3^2+4^2+5^2+6^2 \cdots +9^2}$

Trace → sum of diagonal elements

LU Decomposition
↳ Alan Turing in 1948

$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ $L = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix}$

$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$

$A = LU$

Eg: Given $x_1 + x_2 + x_3 = 1$
$4x_1 + 3x_2 - x_3 = 6$
$3x_1 + 5x_2 = 3x_3 = 4$

$A = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 3 & -1 \\ 3 & 5 & 3 \end{bmatrix}$ $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = C = \begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix}$

$AX = C$

Imp concepts here
Row - echelon form
Gaussian Elimination
Reduce row-echelon form
Rank of a matrix

Eigen values & Eigen vectors.
Vector Operations.

Skewness	Kurtosis
→ shape & size of variation on either side of the central value; extent to which deviation above or below average.	→ freq dist at the central value

Covariance and Correlation.

Covariance

→ Relationship b/w pair of random variables where change in 1 variable causes change in another variable.

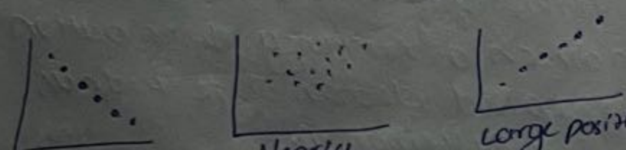
→ Any value b/w -infinity to +infinity

- value → negative relationship

+ value → positive relationship.

→ Used for linear relationship b/w variables

$$\text{Covariance}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)}{n}$$



Large Negative Covariance — Nearly zero covariance — Large positive covariance

Correlation

→ How strongly pairs of variables are related to each other

→ values b/w -1 to +1

+1 → strong positive relation

-1 → strong negative.

$$\text{Corr}(x,y) = \frac{\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\mu_x)^2 \sum_{i=1}^{n}(y_i-\mu_y)^2}}$$

Covariance	Correlation
→ how much 2 variables vary together	→ how strongly 2 variables are related
→ have dimensions	→ dimensionless
→ 2 variables	→ multiple variables

Descriptive Statistics.

Types of Data

- Categorical
- Numerical
  - Discrete — Eg: SAT scores; • Grades; • no of objects
  - Continuous → Infinite values ∞ — Eg: weight; can take all values. • Height • Area • Distance • Time

Levels of Measurement

- Qualitative
  - Nominal → Can't be ordered; Eg: BMW, Mercedez; winter, summer
  - Ordinal → Disgusting, Unappetizing, Neutral, Tasty, Delicious — Ordered
- Quantitative
  - Interval
  - Ratio

Inferential Statistics

→ Methods that involve Probability theory & Distributions to predict population values based on sample data

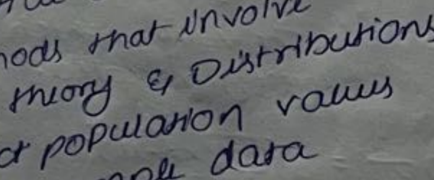
Distribution

A distribution is a function that shows the possible values for a variable & how often they occur.

Probability Dist → eg Dice → $\frac{1}{6}$

Binomial Dist

Normal Dist (Gaussian Dist)



$N \sim (\mu, \sigma^2)$

Statistics.

Population & Sample

Population is collection of all items of interest N

Sample is a subset of the population n

Population is hard to define & hard to analyze

→ Samples have 2 imp features Randomness & Representativeness

Mean, Variance & Standard Deviation & Variance

Mean → Avg. (μ)

Median → $\left(\frac{n+1}{2}\right)^{th}$ value ← odd

$\frac{n/2 + (n/2+1)}{2}$ ← even

Mode → The value which is most commonly repeated.

$SD\ \sigma = \sqrt{\frac{\sum (x_i-\mu)^2}{N}}$

Variance $= \sigma^2$

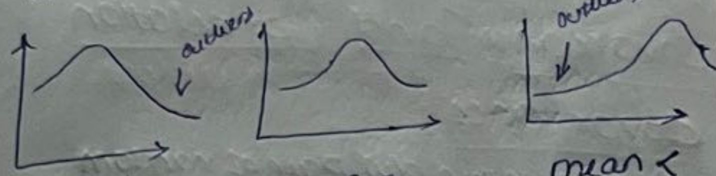
All these are called Measure of central Tendency.

Measures of Asymmetry.

① Skewness

→ Tells abt where the data is situated

Positive Skewness | Zero Skew | Negative Skewness



Right outliers

mean > meadian

mean = median = mode

mean < median

→ Skewness is the link b/w central tendency measures & probablity theory.

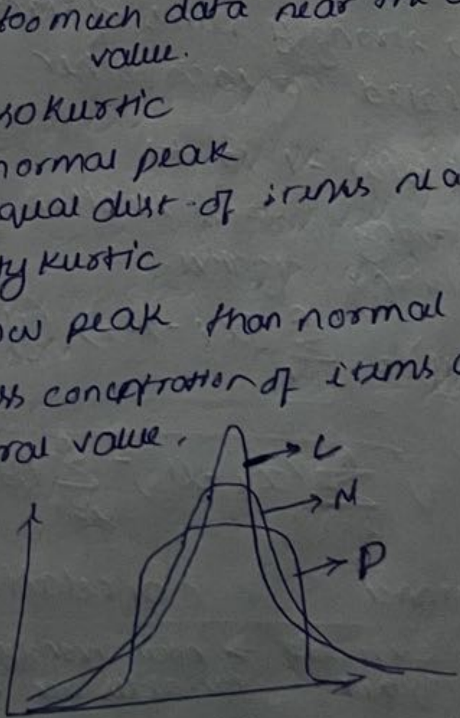
② Kurtosis

Measure of Kurtosis is the extent to which a frequency distribution is peaked, degree of peakedness of a distribution.

a) Leptokurtic distribution.
→ high peak than normal curve
→ too much data near the central value.

b) Mesokurtic
→ normal peak
→ equal dist. of items near central value

c) Platykurtic
→ low peak than normal
→ less concentration of items around central value.



Hypothesis Testing

~~A~~ data is there for do analysis & test

Null Hypothesis H_0 - that we are trying to ~~reject~~ prove reject

alternate hypothesis H_1 → " ~~reject~~ approve

H_0 → Significance Level (α)

The probability of rejecting the null hypothesis, if it is true.

$\alpha = 0.01, 0.05$ or 0.1

Eg: Test if a mc is working properly?

Remaining Topics in Math

Probability Distribution

Regression

Dimensionality Reduction.

Little more on remaining stats topics if time there.

Calculus.

Implicit Differentiation
→ makes use of chain rule of differentiation.

consider eqn of circle

$x^2+y^2=r^2$

$y=\pm\sqrt{r^2-x^2}$

$\frac{dy}{dx}=\frac{1}{2}(r^2x^2)(2x)$

$\frac{dy}{dx}=x(r^2x^2)$

$y=\pm\sqrt{r^2x^2}$ w.k.t $\frac{dy}{dx}=\frac{-x}{y}$.

Revise differentiation & integration resulting answers.

Differentiation

$\frac{d}{dx}x^n=nx^{n-1}$

$\frac{d}{dx}K=0$

$\frac{d}{dx}e^x=e^x$

$\frac{d}{dx}a^x=a^x\log a$

$\frac{d}{dx}(\ln x)=1/x$

$\frac{d}{dx}(\log_a x)=1/x\ln a$

$\frac{d}{dx}\sin x=\cos x$

$\cos x=-\sin x$

$\tan x=\sec^2 x$

$\cot x=-\mathrm{cosec}^2 x$

$\sec x=\sec x\tan x$

$\mathrm{cosec}\, x=-\mathrm{cosec}\, x\cot x$

$\sin^{-1}x=\frac{1}{\sqrt{1-x^2}}$

$\cos^{-1}x=-\frac{1}{\sqrt{1-x^2}}$

$\tan^{-1}x=\frac{1}{1+x^2}$

$\cot^{-1}x=\frac{-1}{1+x^2}$

$\sec^{-1}x=\frac{1}{x\sqrt{x^2-1}}$

Integration

$\int x^n\,dx=x^{n+1}/(n+1)+C$

$\int K\,dx=Kx+C$

$\int e^x\,dx=e^x+C$

$\int a^x dx=a^x/\log a+C$

$\int 1/x\,dx=\ln x+C$

$\int \log_a x\,dx=x\log_a x-x/\ln a$

$\int \sin x\,dx=-\cos x+C$

$\cos x\,dx=\sin x+C$

$\tan x=1/a\ln|\sec x|+C$

$\cot x=1/a\ln|\sin x|+C$

$\sec x=1/a\ln|\sec x+\tan x|+C$

Differentiation	Integration
→ Process of determining the rate of change of a qty w.r.t other qty	→ Process of bringing smaller components into single unit that acts as 1 single component
→ used to find slope of a function	→ Used to find area under the curve
→ Derivatives are considered at a point	→ integrals are considered over an interval
→ $\frac{d}{dx}f(x)$ is unique	→ not be unique since o/p has a C which is arbitrary.