Prove how the given source statement is transformed after each phase of compilation

$x = y + z * 60$

phase 1 - Lexical analysis
" 2 - Syntax " -
" 3 - semantic " - type info
4 - intermediate code generation
5 - code optimization

$x = y + z * 60$

↓

Lexical Analysis

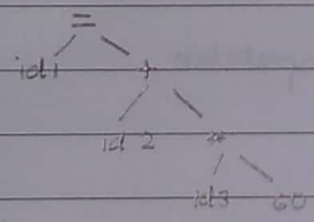
↓

<id 1, x> <op1 = > <id 2, y> <op2, +> <id 3, z> <op3, * > <num, 60 >

↓

Syntax Analysis

↓

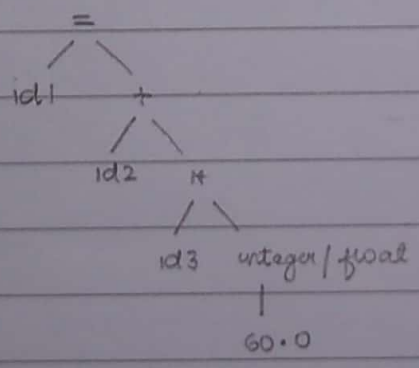


Start → Astart
Astart → id = expr
expr → id + id / id * Num

↓

Semantic Analysis

↓



↓

Intermediate code generation

↓

t_1 = int to float (60)
t_2 = id3 * t_1
t_3 = id2 + t_2
id1 = t_3

↓

Code optimization

t_1 = id3 * 60.0
id1 = id2 + t_2

↓

Code generation

↓

```
MOVF id3, R2
MULF 60.0, R2
MOVF id2, R1
ADDF R2, R1
MOVF R1, id1
```

Symbol table → has identifiers & their types

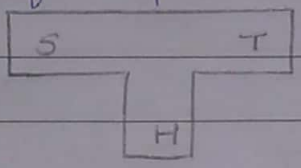
↓

created in 1st phase

but data used in all phases

Bootstrapping and porting

T-diagram of a compiler



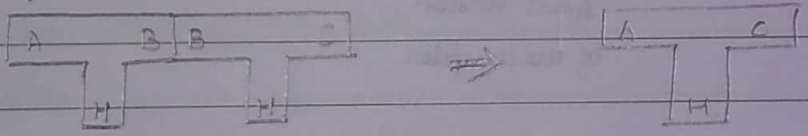
S - source language - that it compiles

T - target " " " generates code for

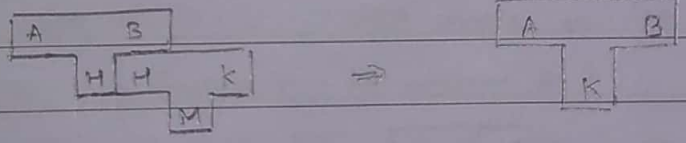
H - implementation language that it is written it

T-diagram can be combined in 2 ways

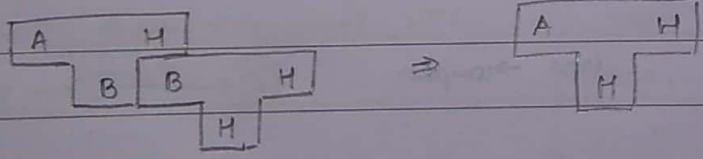
1.



2.

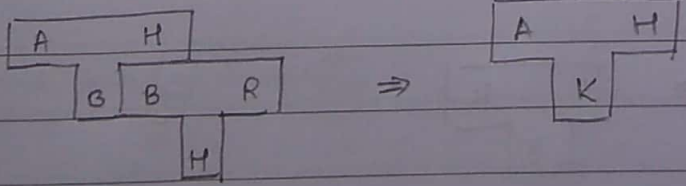


Scenario 1



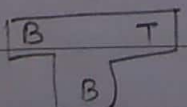
Native compiler

Scenario 2



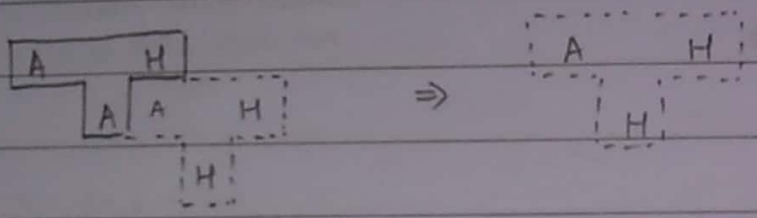
Cross compiler

A ~~compiler~~ compiler written in the same language that is to compile



Circulating problem - No compiler for the source lang yet exists

Step 1

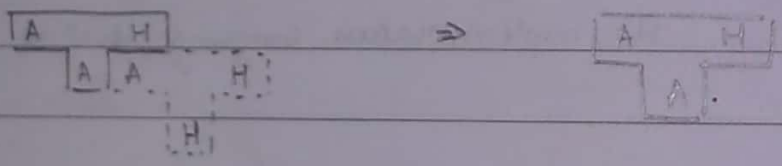


Compiler written in its own language

Quick & dirty compiler written in machine language

Running but inefficient compiler

Step 2

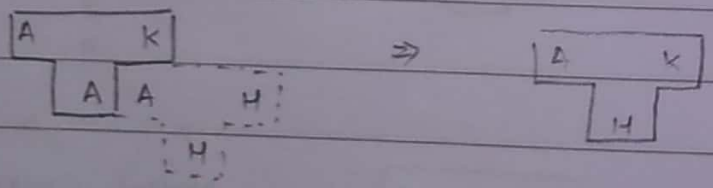


Running but inefficient compiler

final version of the compiler

Porting

Step 1:

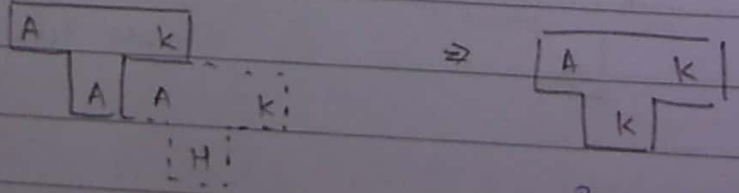


compiler service code ~~retargeted~~ retargeted to k

original compiler

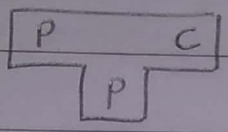
Cross → compiler

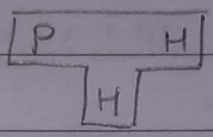
Step 2:



Retargeted compiler

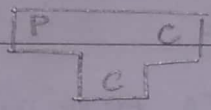
given :



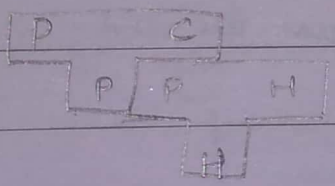


P - pascal

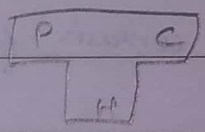
output



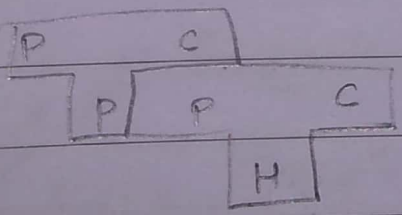
Step 1



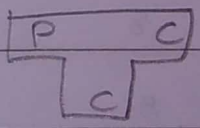
⇒



Step 2



⇒



offline

UNIT 2

$A \to A\alpha / \beta$ left recursive

$A \to \alpha A / \beta$ right recursion

Elimination of left recursion

(i) $A \to A\alpha / \beta$

replace with

$A \to \beta A'$

$A' \to \alpha A' / \varepsilon$

(ii) $A \to A\alpha_1 / \cdots / A\alpha_m / \beta_1 / \cdots / \beta_n$

replace with

$A \to \beta_1 A' / \cdots / \beta_n A'$

$A' \to \alpha A' / \cdots / \alpha_m A' / \varepsilon$

(iii) input : Grammar G with no cycles or ε productions

Output : An equivalent grammar with no left recursion

method :

① arrange the NTs in order $A_1, A_2 \cdots A_n$

② for i = 1 to n do begin

for j = 1 to i-1 do begin

replace each production of the form $A_i \to A_j \gamma$ by the productions $A_i \to \delta_1 \gamma / \cdots / \delta_k \gamma$ where $A_j \to \delta_1 / \cdots / \delta_k$ are all current A_j productions and eliminate the immediate left recursion among the A_i productions

end

egs :

eg: $E \to E+T / T$ ($A \; A\alpha \; \beta$)

$T \to T*F / F$ ($A \; A\alpha \; \beta$)

$F \to (E) / id$ → no cycle

G' $E \to TE'$

$E' \to +TE' / E$

$T \to FT'$

$T' \to *FT' / E$

$F \to (E) / id$

ii G: ① $S \to AB$

② $A \to CB \mid b$

③ $C \to Sa$

④ $B \to a$

$S \to AB \to \underline{CB}B \to SaBB$

$S \to SaBB$ ∴ indirect left recursion

Indirect left recursion

$n = 4$

$i = 1$ $j = 1$ $A_1 \to A_1$ $S \to S$ No

$i = 2$ $j = 1$ $A_2 \to A_1$ $A \to S$ No

$i = 3$ $j = 1$ $A_3 \to A_1$ $C \to S$ Yes

$C \to ABa$ [$B \to AB$]

$j = 2$ $A_3 \to A_2$ $C \to A$ Yes

$C \to CBBa \mid bBa$

$A \to CB$ replace in $C \to ABa$

Eliminate left recursion from $C \to CBBa \mid bBc$

$C \to bBa\,C'$

$C \to BBa\,C' \mid E$

$i = 4$ $j = 1$ to 3 $B \to S$ $B \to A$ $C \to C$ No

grammar becomes

G': $S \to AB$

$A \to CBb$

$C \to bBa\,C$

$C' \to BBd\,C' \mid E$

$B \to a$

G : $S \to (L) \mid a$

$L \to L, S \mid S$

(i) (a, a)

$S \underset{lm}{\Rightarrow} (L)$

$\underset{lm}{\Rightarrow} (L, S)$

$\underset{lm}{\Rightarrow} (S, S)$

$\underset{lm}{\Rightarrow} (a, S)$

$\underset{lm}{\Rightarrow} (a, a)$

$S \underset{rm}{\Rightarrow} (L)$

$\underset{rm}{\Rightarrow} (L, S)$

$\underset{rm}{\Rightarrow} (L, a)$

$\underset{rm}{\Rightarrow} (S, a)$

$\underset{rm}{\Rightarrow} (a, a)$

ii (a, (a, a))

Left factoring

$A \rightarrow \alpha\beta_1 / \alpha\beta_2 / \gamma$

Replace with

$A \rightarrow \alpha A' / \gamma$

$A' \rightarrow \beta_1 / \beta_2$

eg G : $A \rightarrow qB / qC$

$\alpha\beta_1 \quad \alpha\beta_2$

$A \rightarrow qA'$

$A' \rightarrow B / C$

Eg 2

G : stat → pQ | pR | ε

Q → b stat | pQQ

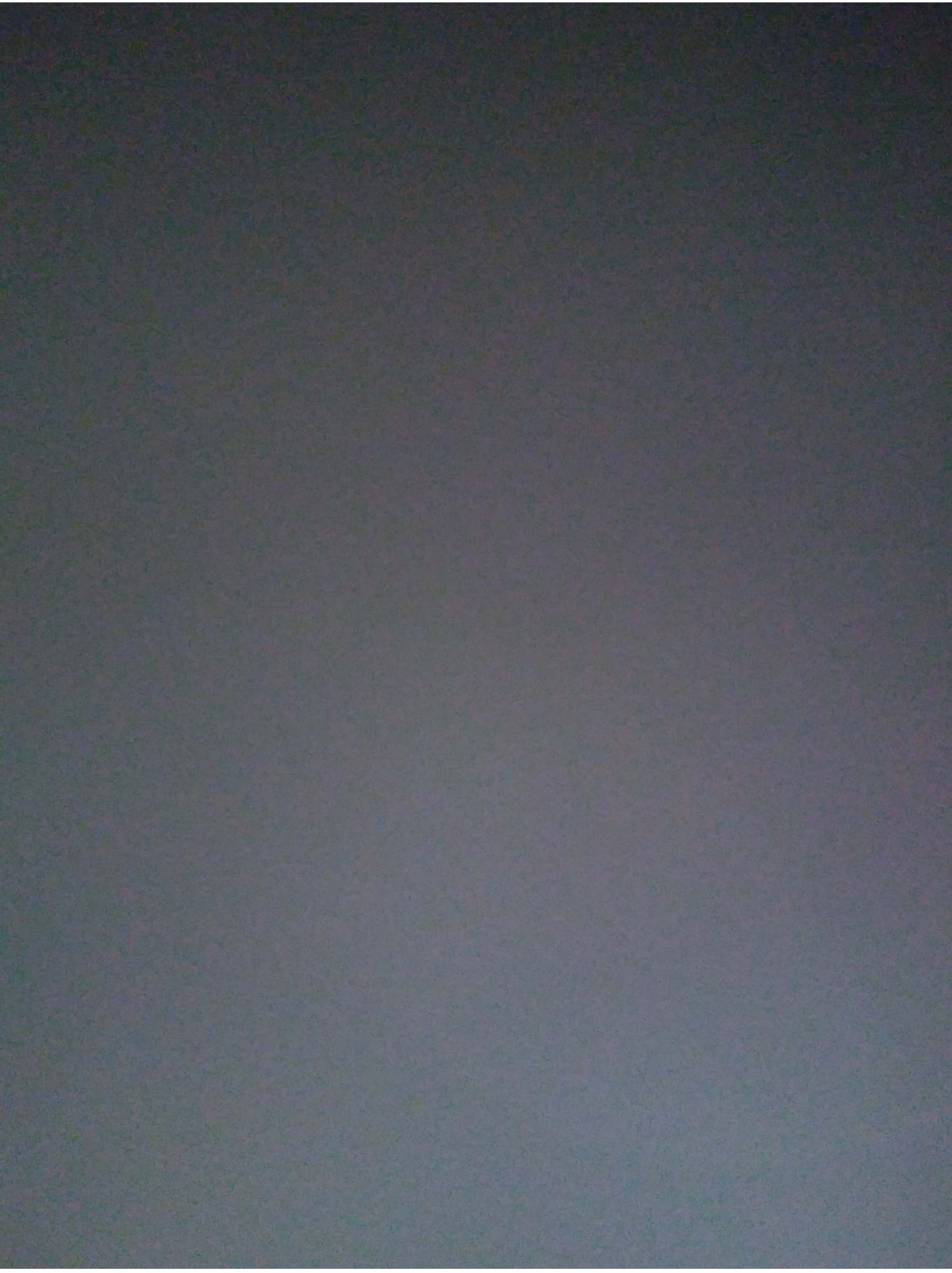
R → p stat | bRR

G' : stat → p stat' | ε

stat' → Q | R

Q → b stat | pQQ

R → p stat | bRR



27 observation

Q For full data for attribute age 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70 sketch egs of each of following sampling techniques SRSWOR, SRSWR, cluster sampling, Stratified sampling. Use sample size of 5 and the strata: Youth, middle age, Senior

Youth | Middle age | Senior (above 60)

T_1	13	T_{10}	25	T_{19}	35
T_2	15	T_{11}	25	T_{20}	35
T_3	16	T_{12}	25	T_{21}	35
T_4	16	T_{13}	25	T_{22}	36
T_5	19	T_{14}	25	T_{23}	40
T_6	20	T_{15}	30	T_{24}	45
T_7	20	T_{16}	33	T_{25}	46
T_8	21	T_{17}	33	T_{26}	52
T_9	22	T_{18}	33	T_{27}	70

SRSWOR

Noise - random error or a variance in a measured variable

Eg Attribute values are numeric Eg attribute like price of product
how can v remove this noise there r some data smoothing techniques

① Binning - it smooths a sorted data value by consulting its neighbourhood " value around it. They perform local smooth

- Smoothing by bin means - each value in bin is replaced by the mean value of the bin
- " " " median - replaced by median value of the bin
- " " " boundaries - identify the min & max of each bin

Each value is replaced by the closest boundary value.

② Regression - conforms data value to a funct^n there r 2 types

- linear
- multiple linear regression.

Problem

Suppose that the data for analysis includes the attribute age the age values for the data tuples are (in ↑ing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

(a) Use smoothing by bin means to smooth the above data by using bin depth of 3 & then comment on effect of this technique for given data

- Sup 1 sort data
- sup 2: Partition data into depth = 3

Bin 1 : 13, 15, 16 Bin 4 : 22, 25, 25 Bin 7 : 35, 35, 35

Bin 2 : 16, 19, 20 5 : 25, 29, 30 8 : 36, 40, 45

Bin 3 : 20, 21, 22 6 : 33, 33, 35 9 : 46, 52, 70

- Calculate arithmetic mean of each bin

14.667 ; 18.333 ; 21 ; 24 ; 26.667 ; 33.667 ; 35 ; 40.33 ; 56
Bin 1 Bin 2 Bin 3 Bin 4 Bin 5 Bin 6 7 8 9

- Replace each of values in each bin by arithmetic mean

Bin 1 : 14.667, 14.667, 14.667 B4 : 24, 24, 24 7 : 35, 35, 35

Bin 2 : 18.33, 18.33, 18.33 Bin 5 : 26.667, 26.667, 26.667 8 : 40.33, 40.33, 40.33

Bin 3 : 21, 21, 21 6 : 33.667, 33.667, 33.667 9 : 56, 56, 56

b use smoothing by bin boundaries using bin depth of 3

step 1 & 2 same as prev

Bin 1: 13, 16, 16 4: 22, 25, 25 7: 35, 35, 35

2: 16, 20, 20 5: 25, 25, 30 8: 36, 36, 45

3: 20, 20, 22 or 20, 22, 22 6: 33, 33, 35 9: 46, 46, 70

c. Wat other methods are there for data smoothing

Data Preprocessing

data transformation

data are transformed or consolidated into forms appropriate for mining

Strategies

1 Smoothing - works to remove noise from the data

- Binning
- Regression
- clustering

2 . . .

3 Aggregation - summary or aggregation operations are applied to the data

4. Normalization - attribute data are scaled so as to fall within a smaller range such as -1.0 to 1.0 or 0.0 to 1.0

5. Discretization - raw values of a numeric attribute (eg age) are replaced by interval labels (eg 0-10, 11-20 ....) or conceptual labels (eg youth, adult or senior) the labels can be recursively organized into higher level concepts resulting in a concept hierarchy for numeric attribute

6. Concept hierarchy generation for nominal data - attributes such as street can be generalized to higher level concept like city or country

change frm kg to pounds will give diff results
attribute with higher value will have higher weight

Data transformation by Normalizatn

- this helps avoid dependence on choice of measurements units
- normalization gives all attributes equal weight
- useful for classification algorithm & clustering
- There r many methods for ~~two~~ normalizatn

- ① Min-max algo

this performs a linear transformation on original data

Suppose min_A and max_A are minimum & maximum values of an attribute. Min-max normalizatn maps a value v_i of A to v_i' in the range [new_min A, new_max A] by computing

$$v_i' = \frac{v_i - min A}{max_A - min_A} (new\text{-}max_A - new\_min A) + new\_min_A$$

income has min 12000 Rs

max 98000 Rs

we would like to map income in range [0.0, 1.0]

What will 73,600 Rs for income transformed to using min max normalizatn

$$v_i' = \frac{v_i - min A}{max_A - min A}$$

$$= \frac{73600 - 12000}{98000 - 12000} (1.0 - 0.0) + 0$$

$$= \frac{61600}{86000}$$

$$v_i' = 0.716$$

② z-score normalization (zero mean normalization

the values for an attribute A are normalized based on the mean (ie average) & the standard deviation of A

The value v_i of A is normalized to v_i' by computing $v_i' = \frac{v_i - \bar{A}}{\sigma_A}$

$\bar{A}$ = mean & σ_A = standard deviatn of attribute A

$$\bar{A} = \frac{1}{n}(V_1 + V_2 + \cdots + V_n) \qquad \sigma_A = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Eg Suppose the mean & SD of the values of income are 54000Rs & 16000Rs resp what is the value for the income 73600 using zscore normalizatn

$$V_i' = \frac{73600 - 54000}{16000} = 1.225$$

③ Normalization by decimal scaling

this normalizes by moving decimal point of values of attribute A the no of decimal points moved depends on the maximum absolute value of A. A value V_i of A is normalized to V_i' by computing $V_i' = \frac{V_i}{10^j}$ where j = smallest integer such that ~~maximum of~~ $\max(|V_i'|) < 1$

Eg suppose that the recorded values of A range from -986 to 917 the max absolute value of A is 986 to normalize by decimal scaling v divide each value by 1000 (ie j = 3) so that -986 normalizes to -0.986 & 917 normalizes to 0.917

Q Normalise following group of data using (i) min max normalizatn by setting min = 0 & max = 1 ii z score normalizatn range 0.0 - 1.0

data : 200, 300, 400, 600, 1000 min = 200 max 1000

i $V_i' = \frac{V_i - \min A}{\max_A - \min_A}(\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A$

200 $V_i' = \frac{200 - 200}{1000 - 200}(1.0 - 0.0) + 0.0 = 0$

400 $\frac{400 - 200}{1000 - 200} = \frac{200}{800} = 0.25$

300 $\frac{300 - 200}{1000 - 200} = \frac{100}{800} = 0.125$

600 $\frac{600 - 200}{1000 - 200} = \frac{400}{800} = 0.5$

1000 $\frac{1000 - 200}{1000 - 200} = \frac{800}{800} = 1$

ii 200

$$\bar{A} = \frac{1}{5}(200 + 300 + 400 + 600 + 1000)$$

$$= 500$$

$$\sigma_A = \sqrt{\frac{1}{5}\sum_{i=1}^{5}(x_i - \bar{x})^2}$$

$$= \frac{1}{5}\{(200-500)^2 + (300-500)^2 + (400-500)^2 + (600-500)^2 + (1000-500)^2\}$$

$90000 \quad 40000 \quad 10000 \quad 10000 \quad 250000$

$$= 282.8$$

$$v_i' = \frac{v_i - \bar{A}}{\sigma_A}$$

200 $v_i' = -1.06$

300 $= -0.7$

400 $= -0.35$

600 $= 0.35$

1000 $= 1.78$

Discretization and Binarization

transform a continuous attribute into a categorical attribute (descretizatn) and both continuous & discrete attributes need to be transformed into 1 or more binary attributes (binarization)

Binarizatn

A simple technique of Binarizatn of a category will be as follows if there are m categorical values then uniquely assign each original value to the integer in the interval $[0, m-1]$ then convert each of these m integers to a binary no. Since $n = \lceil \log_2 m \rceil$ binary digits are required to represent the integers the binary nos are represented using n binary attributes

Eg categorical variable with 5 values {awful, poor, ok, good, great}

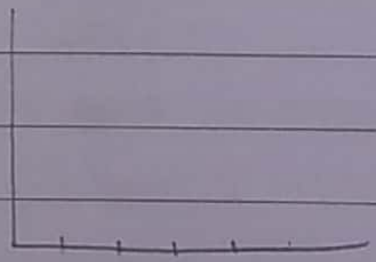
∴ there r 5 values v would require 3 binary variables x_1, x_2, x_3 & conversion is shown in table

attribute type	dissimilarity	similarity
Nominal	$d = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$	$s = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$
Ordinal	$d = \lvert x - y \rvert / (n-1)$ (values mapped to integers 0 to $n-1$, where n = no. of values)	$s = 1 - d$
Interval ratio	$d = \lvert x - y \rvert$	$s = -d$ $\quad s = \frac{1}{1+d}$ $\quad s = e^{-d}$

$$s = \frac{1 - d - min\_d}{max\_d - min\_d}$$

Euclidean Distance

$$d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$



$P_1 = (0, 2)$ $\quad P_2 = (2, 0)$ $\quad P_3 = (3, 1)$ $\quad P_4 = (5, 1)$

$P_2 P_3 = \sqrt{(-1)^2 + (1)^2} = \sqrt{2} = 1.414$

$P_1 P_3 = \sqrt{(-3)^2 + (1)^2} = \sqrt{10}$

$P_1 P_4 = 25 + 1 = \sqrt{26} = 5.099$

$P_2 P_1 = (2)^2 + (-2)^2 = \sqrt{8}$

	p_1	p_2	p_3	p_4
p_1	0	2.828	3.162	5.099
p_2	2.828	0	1.414	3.162
p_3	3.162	1.414	0	2
p_4	5.099	3.162	2	0

$P_2 P_4 = \sqrt{4 + 1} = 3.162$

$P_3 P_4 = \sqrt{2^2} = \sqrt{4} = 2$

$d(p_1, p_2) = \sqrt{(0-2)^2 + (2-0)^2}$

$= \sqrt{2^2 + 2^2} = \sqrt{8}$

$= 2.828$

$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

$d(p_1, p_2) = d(p_2, p_1)$ $\quad d(p_1, p_1) = 0$

Minkowski Distance - generalizatn of euclid

$$d(x,y) = \left(\sum_{k=1}^{n} |x_k - y_k|^r \right)^{1/r}$$

Eg $P_1(0,2)$ $P_2(2,0)$ $P_3(3,1)$ $P_4(5,1)$

L1 $r = 1$

	P_1	P_2	P_3	P_4
P_1	0	4	4	6
P_2	4	0	2	4
P_3	4	2	0	2
P_4	6	4	2	0

L2 $r = 2$

	P_1	P_2	P_3	P_4
P_1	0	2.828	3.162	5.099
P_2	2.828	0	1.414	3.162
P_3	3.162	1.4414	0	2
P_4	5.099	3.162	2	0

L∞ $r = \infty$

	P_1	P_2	P_3	P_4
P_1	0	2	3	5
P_2	2	0	1	3
P_3	3	1	0	2
P_4	5	3	2	0

Common properties of distance

① $d(x,y) \geq 0$ $\forall x$ and y and $d(x,y) = 0$ iff $x = y$

② $d(x,y) = d(y,x)$ $\forall x$ and y (symmetry)

③ $d(x,z) \leq d(x,y) + d(y,z)$ $\forall$ points x, y and z (Δle inequality)

where $d(x,y)$ = distance (dissimilarity) betⁿ points (data objects) x and y

Similarity betⁿ binary vector

patient table

name	gender	fever	cough	test 1	test 2	test 3	test 4
Jack	M	Y	Y	P	N	P	N
John	M	Y	N	N	N	N	P
Mary	F	Y	Y	P	N	P	P
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(these are asymmetric attributes)

Simple Matching and Jacard coefficients

$SMC = \frac{\text{no. of matches}}{\text{no of attributes}}$

$= \frac{(f_{11} + f_{00})}{(f_{01} + f_{10} + f_{11} + f_{00})}$

f_{01} = no of attributes where $x=0$ $y=1$

f_{10} = $x=1$ $y=0$

f_{00} = $x=y=0$

f_{11} = $x=y=0$

$J = \frac{\text{no of 11 matches}}{\text{no. of non 0 attributes}}$

$= \frac{f_{11}}{f_{01} + f_{10} + f_{11}}$

- when u have assymetric J plays an imp role
- when u have symmetric SMC plays an imp role

Jacard coefficient (Jack, Jim) assume yes = 1 (Y) no = 0 (N) +ve P=1 -ve N=0

$\frac{1}{1+3+1} = \frac{1}{5} = 0.2$

Cosine Similarity

Document Vector

document	team	coach	hockey	baseball	soccer	penalty	score	win	loss	season
doc 1	5	0	3	0	2	0	0	2	0	0
doc 2	3	0	2	0	1	1	0	1	0	1
doc 3	0	7	0	2	1	0	0	3	0	0
doc 4	0	1	0	0	1	2	2	0	3	0

term frequency vector of "5 0 3 0 2 0 0 2 0 0" of is term freq vector of doc 1

let x & y be 2 vectors for comparison

similarity is given by

$$Sim(x,y) = \frac{x \cdot y}{\|x\| \|y\|}$$

where $\|x\|$ = euclidean norm of vector $x = (x_1, x_2 \dots x_n)$ defined as $\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ conceptually its the length of vector

similarly $\|y\|$ is the euclidean norm of vector y

compute the cosine of the angle between x & y

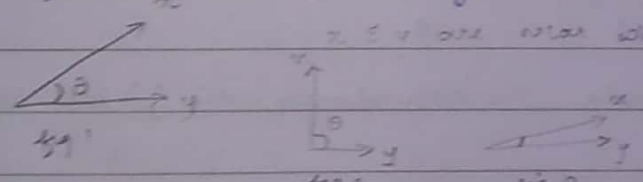


fig 1

fig 2

fig 3

- similarity bet doc 1 and doc 2

x = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0) = term freq of doc 1

y = (3, 0, 2, 0, 1, 1, 0, 1, 0, 1) = " " " 2

$x \cdot y = 5\times3 + 0\times0 + 3\times2 + 0\times0 + 2\times1 + 0\times1 + 0\times0 + 2\times1 + 0\times0 + 0\times1$

$= 15 + 6 + 2 + 2 = 25$

$\|x\| = \sqrt{25+0+9+0+4+0+0+4+0+0}$

$= \sqrt{25+9+4+4} = \sqrt{42} = 6.48$

$\|y\| = \sqrt{9+0+4+0+1+1+0+1+0+1}$

$= \sqrt{9+4+4} = \sqrt{17} = 4.12$

$$Sim(x, y) = \frac{25}{6.48 * 4.12} = \frac{25}{26.697} = 0.94$$ very close to 1 ∴ polar

Sim (doc 1, doc 4)

$x = (5, 0, 3, 0, 2, 0, 0, 2, 0, 0)$

$y = (0, 1, 0, 0, 1, 2, 2, 0, 3, 0)$

$x \cdot y = 2$

$\|x\| = 6.48$

$\|y\| = \sqrt{1+1+4+4+9} = \sqrt{19} = 4.358$

$$Sim(x, y) = \frac{2}{6.48 \times 4.358} = \frac{2}{28.239} = 0.0708$$

Dissimilarity for attributes of mixed type

Object identifier	Test 1 (nominal)	Test 2 (Ordinal)	Test 3 (numeric)
1	Code A	Excellent	45
2	Code B	fair	22
3	Code C	good	64
4	Code A	excellent	28

In such situat^n v use the formula. The dissimilarity $d(i, j)$ bet^n objects i & j is defined as

$$d(i, j) = \frac{\sum_{f=1}^{p} \delta_{ij}^{(f)} \cdot d_{ij}^{(f)}}{\sum_{f=1}^{p} \delta_{ij}^{(f)}}$$

where we $\delta_{if}^{(f)} = 0$ if either :

x_{if} or x_{jf} is missing (ie there is no measurement of attribute f for object i or object j)

OR $x_{if} = x_{jf} = 0$ the attribute f is assymmetric binary otherwise $\delta_{if}^{(f)} = 1$

the contribut^n of attribute f to the dissimilarity between i & j ie $d_{ij}^{(f)}$ is computed depending on its type

① if f is numeric $d_{ij}^{(f)}$ is computed using $d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max x_{hf} - \min x_{hf}}$

where h runs over all non missing objects for attribute f

② if f is nominal or binary $d_{ij}^{(f)} = 0$ if $x_{if} = x_{jf}$

$d_{ij}^{(f)} = 1$ otherwise

③ if f = ordinal compute the ranks r_{if} & $z_{if} = \frac{r_{if} - 1}{M_f - 1}$ & treat z_{if} as numeric

No. of values

dissimilarity matrix for test 3 numeric

	1	2	3	4
1	0			
2	0.55	0		
3	0.45	1	0	
4	0.40	0.14	0.86	0

dissimilarity for test 2

	1	2	3	4
1	0			
2	1	0		
3	0.5	0.5	0	
4	0	1	0.5	0

fair - 0
good - 1
excellent - 2

dissimilarity for test 1

	1	2	3	4
1	0			
2	1	0		
3	1	1	0	
4	0	1	1	0

code B, A

code A & code A

dissimilarity	1	2	3	4
1	0			
2	0.85	0		
3	0.65	0.83	0	
4	0.13	0.71	0.79	0

d(4,1) = 0.4/3 d(4,2) 0.14+2 = 0.2:

Objects 1 & 4 r most nlar

" 2 & 1 r " dis-similar

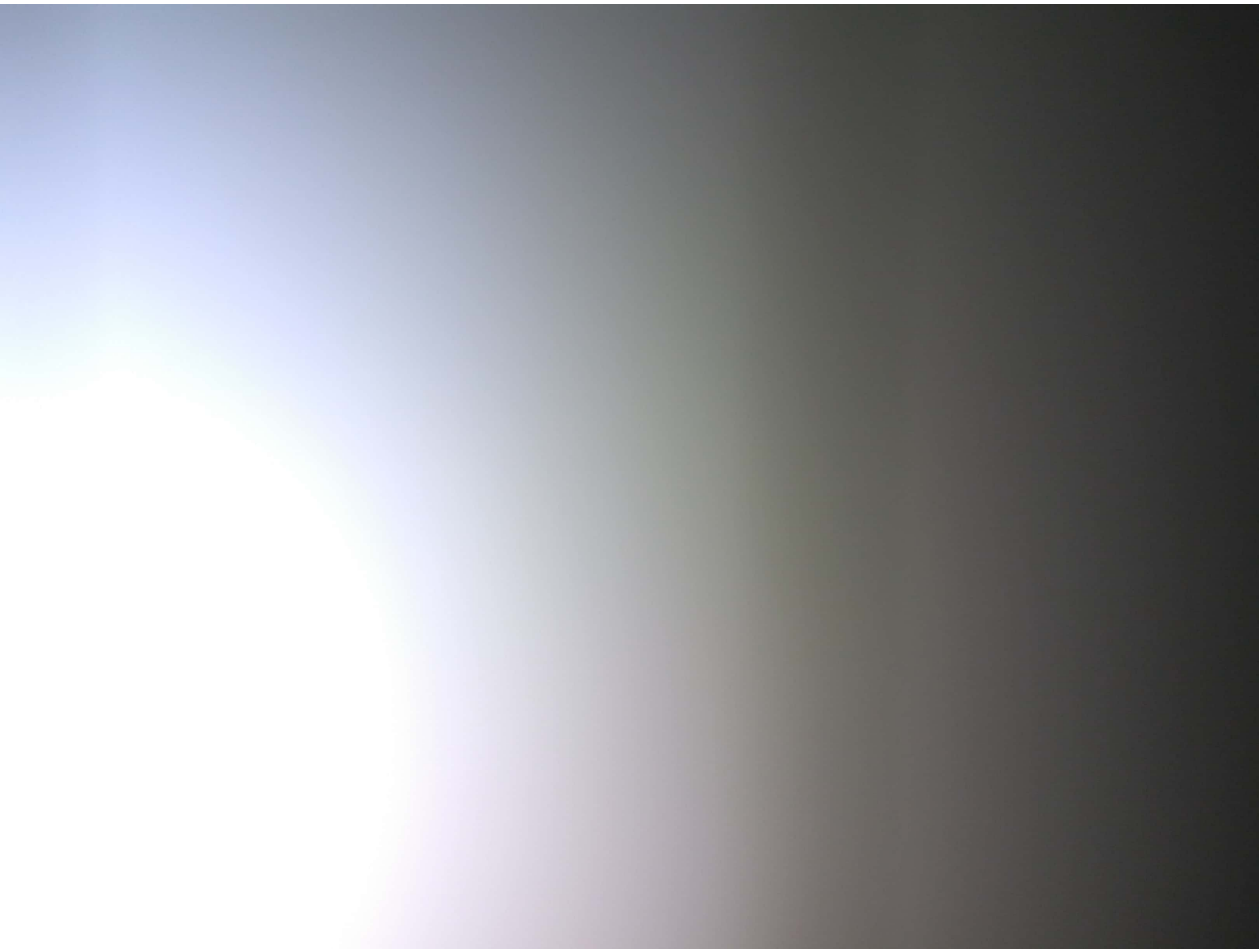
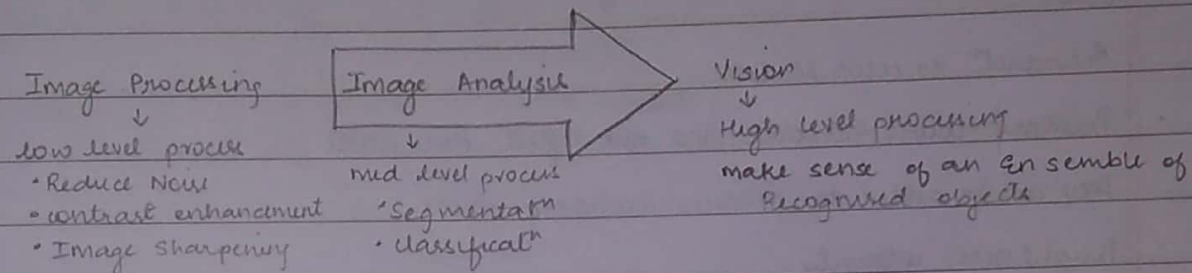


Image processing

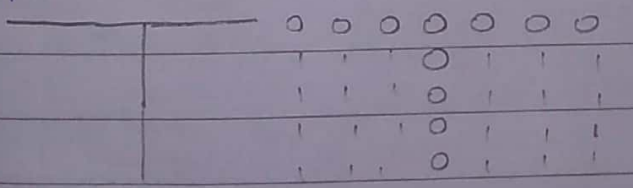


Application

- Medical Images
- Face Recognition
- Finger print Detection
- Pattern Recognition
- Feature Extraction

Defintions :-

- Image - An image is a 2-D signal analog or digital that contains intensity or colour information arranged along the x and y spatial axis
It is a 2D function $f(x, y)$ where x, y are spatial coordinates f is amplitude of a function also called as the intensity or grey level or the color of the image at that point



0 - represents black
1 - " white
these are called a monochrome scheme

- Amplitude - measure of how far and in what direction a variable differs from 0
- Image Processing - IP is a form of signal processing for which the input is an image such as photograph or frames or video and o/p can be an image or set of characteristics or parameters related to the image
- y image processing or y do v need it → is a method of performing operatn on an image to get an enhanced image or to extract useful info from it

- digital image processing is a subset of digital signal processing
- it allows wider range of algo to be applied to the i/p data which can avoid problems

of noise & distortion (no image sharpening)

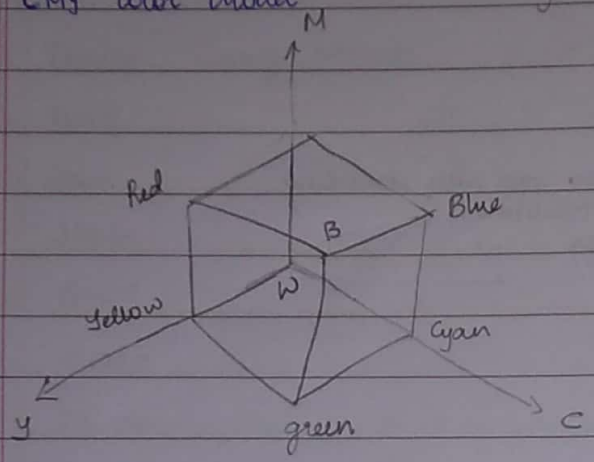
Color

Introduction to color Models

- Primary colors - absorbs and also reflect other colors
  they can be added to produce secondary colors
- Brightness - intensity
- ~~Hue~~ Hue - is a dominant wavelength (color)
- Saturation - amnt of white light mixed with the hue
- Chromaticity - depends on a combination of hue and saturation
- Luminosity - ~~maximum~~ measure of light energy emitted (More brighter - more energy - more luminosity)
- Luminance - measure of luminous intensity per unit area of light
  effuse

Offline - lecture 2

CMY color Model C - Cyan M - Magenta Y - yellow



- this model is used for describing the color o/p to hardcopy devices
- colors are specified by what is subtracted from white color

  Green + Blue = Cyan

- when white light is reflected from cyan color mean the reflected light doesn't have the red component ie red light is absorbed or subtracted

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$

$$I_{CMY} = (l_c \quad l_m \quad l_y)$$

$$l_c = 1 - \frac{f_R(x,y)}{f_R(x,y) + f_G(x,y) + f_B(x,y)}$$

$$l_m = 1 - \frac{f_G(x,y)}{f_R(x,y) + f_G(x,y) + f_B(x,y)}$$

$$l_y = 1 - \frac{f_B(x,y)}{f_R(x,y) + f_G(x,y) + f_B(x,y)}$$

Convert RGB values to CMY model

RGB values (29, 98, 128)

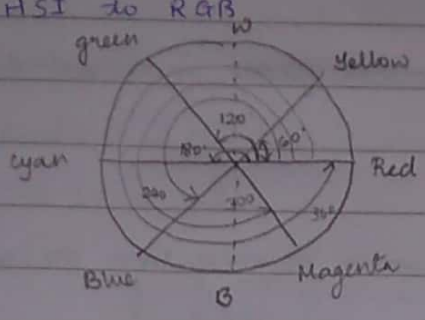
$$\frac{29}{255} = 0.113 \qquad \frac{98}{255} = 0.384 \qquad \frac{128}{255} = 0.5019$$

Cyan = 1 − R
= 1 − 0.113
= 0.887

Magenta = 1 − G
= 1 − 0.384
= 0.616

Yellow = 1 − B
= 1 − 0.501
= 0.499

HSI to RGB



angles r = clock Red = 0°
Yellow = 60° green = 120° cyan = 180° Blue = 240°
Magenta 300° Red = 360

Note H = ?

H = 0 at H = 0 its red color

$R = I + 2IS$ *

$G = I - IS$

$B = I - IS$

if $0 < H < 120$ Blue not present ∴ I - IS

$R = I + IS \times \cos(H) / \cos(60 - H)$

$G = I + IS \times [1 - \cos(H) / \cos(60 - H)]$

$B = I - IS$

H = 120° its green

$R = I - IS$

$G = I + 2IS$ *

$B = I - IS$

if $120 < H < 240$ Red not present

$R = I - IS$

$G = I + IS \times [\cos(H - 120) / \cos(180 - H)]$

$B = I + IS\ [1 - \cos(H - 120) / \cos(180 - H)]$

H = 240

$R = I - IS$

$G = I - IS$

$B = I + 2IS$ *

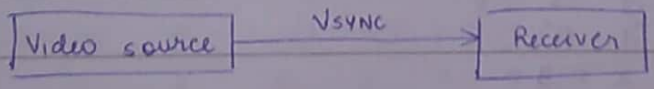
if $240 < H < 360$

$R = I + IS \times [1 - \cos(H - 240) / \cos(300 - H)]$

$G = I - IS$

$B = I + IS \times [\cos(H - 240) / \cos(300 - H)]$

Video



① VSYNC generation
② 1st line image is scanned
③ level synchronisation signal

- Its a sequence of continuous picture and each picture in the sequence is called as a frame

- 25 frames or more per second ~~or more case is good~~ must be displayed

→ term used to compare 2 signals
correlation
is
a process of moving a filter mass over the image in computing the sum of the product at each location

let I = {0, 0, 1, 0, 0} mask = {3, 2, 8}

0 0 1 0 0 ⇒ 0 0 0 0 1 0 0 0 0

3 2 8 3 8 2 3 2 8 3 2 8

(Step i) a 0 padding process 3 2 8 | 0 0 0 0 1 0 0 0 0 | 3 2 8 → step i

ii initial position step

(step iii) right shift by 1 bit position

iii
3 2 8
0 0 0 0 1 0 0 0 0
0 0

iv right shift by 2 bit
3 2 8
0 0 0 0 1 0 0 0 0
0 0 8

v right shift by 3 bit
3 2 8
0 0 0 0 1 0 0 0 0
0 0 8 2

vi right shift by 4 bit
3 2 8
0 0 0 0 1 0 0 0 0
0 0 8 2 3

vii right shift by 5 bit
3 2 8
0 0 0 0 1 0 0 0 0
0 0 8 2 3 0

viii right shift by 6 bit
3 2 8
0 0 0 0 1 0 0 0 0
0 0 8 2 3 0 0

Convolution

I = {0, 0, 1, 0, 0}

K = {3, 2, 8}

step i 0 padding process

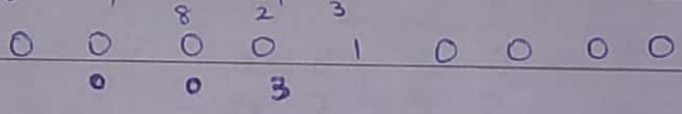
Rotate Mask = 180°

K = {8, 2, 3}

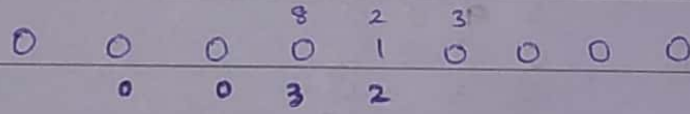
ii mutal slip

iii right shift by 1 position

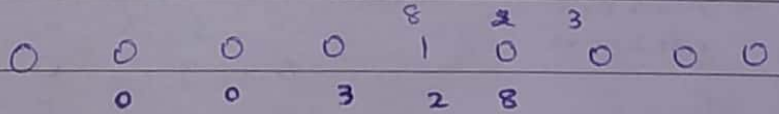
iv right shift by 2 position



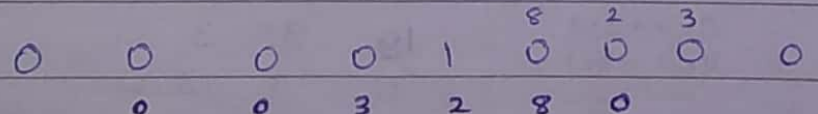
v



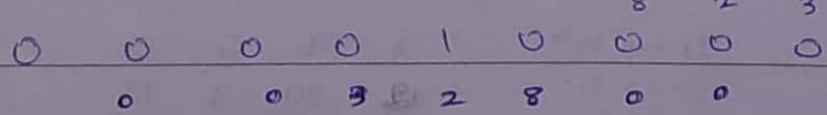
vi



vii



vii



2 D

$I = \begin{Bmatrix} 3 & 3 \\ 3 & 3 \end{Bmatrix}$ be an image & $K = \begin{Bmatrix} 1 & 2 \\ 3 & 4 \end{Bmatrix}$

be a kernel {mask} Perform convolution & correlation

Convolution

① Rotate by 180°

① $\begin{Bmatrix} 1 & 2 \\ 3 & 4 \end{Bmatrix}$ rotate by 180 ⇒ ② $\begin{Bmatrix} 3 \leftrightarrow 4 \\ 1 \leftrightarrow 2 \end{Bmatrix}$ ⇒ ③ $\begin{Bmatrix} 4 & 3 \\ 2 & 1 \end{Bmatrix}$

(a)

④ $\begin{Bmatrix} 3 & 3 \\ 3 & 3 \end{Bmatrix}$ $\boxed{\begin{matrix} 4 & 3 \\ 2 & 1 \end{matrix}}$ kernel

⑤
$$\begin{matrix} \boxed{0^4 \;\; 0^3} & 0 & 0 \\ \boxed{0^2 \;\; 3^1} & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

⑥ $(4\times0)+(3\times0)+(2\times0)+(3\times1) = 3$ o/p written in top left of mask

⑦
$$\begin{matrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(b)

⑧
$$\begin{matrix} 3 & \boxed{0^4 \;\; 0^3} & 0 \\ 0 & \boxed{3^2 \;\; 3^1} & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

⑨ $4\times0 + 0\times3 + 3\times2 + 3\times1 = 9$

⑩
$$\begin{matrix} 3 & 9 & 0 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(c)

⑪
$$\begin{matrix} 3 & 9 & \boxed{0^4 \;\; 0^3} \\ 0 & 3 & \boxed{3^2 \;\; 0^1} \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

⑫ $0\times4 + 0\times3 + 3\times2 + 0\times1$

⑬
$$\begin{matrix} 3 & 9 & 6 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(d)

⑭
$$\begin{matrix} 3 & 9 & 6 & 0 \\ \boxed{0^4 \;\; 3^3} & 3 & 0 \\ \boxed{0^2 \;\; 3^1} & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

$0\times4 + 3\times3$
$0\times2 + 3\times1$

$$\begin{matrix} 3 & 9 & 6 & 0 \\ 12 & 3 & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(e)

$$\begin{matrix} 3 & 9 & 6 & 0 \\ 12 & \boxed{3^4 \;\; 3^3} & 0 \\ 0 & \boxed{3^2 \;\; 3^1} & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

$3\times4 + 3\times3$
$3\times2 + 3\times1$
$12+9+6+3$
30

$$\begin{matrix} 3 & 9 & 6 & 0 \\ 12 & 30 & 3 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(f)

$$\begin{matrix} 3 & 9 & 6 & 0 \\ 12 & 30 & \boxed{3^4 \;\; 0^3} \\ 0 & 3 & \boxed{3^2 \;\; 0^1} \\ 0 & 0 & 0 & 0 \end{matrix}$$

$3\times4 + 0\times3$
$3\times2 + 0\times1$
18

$$\begin{matrix} 3 & 9 & 6 & 0 \\ 12 & 30 & 18 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$$

(g)

3	9	6	0
12	30	18	0
0⁴	3³	3	0
0²	0¹	0	0

$0\times4 + 3\times3 + 0\times2 + 0\times1$

3	9	6	0
12	30	18	0
9	3	3	0
0	0	0	0

(h)

3	9	6	0
12	30	18	0
9	3⁴	3³	0
0	0²	0¹	0

$3\times4 + 3\times3$
$0\times2 + 0\times1$

3	9	6	0
12	30	18	0
9	21	3	0
0	0	0	0

(i)

3	9	6	0
12	30	18	0
9	21	3⁴	0³
0	0	0²	0¹

$3\times4 + 3\times0$
$+0\times2 + 0\times1$

3	9	6	0
12	30	18	0
9	21	12	0
0	0	0	0

What is the highest value ? 30 ↑ have placed mark on the area which contains no

Correlation $\begin{Bmatrix} 1 & 2 \\ 3 & 4 \end{Bmatrix}$

0₁	0₂	0	0
0₃	3₄	3	0
0	3	3	0
0	0	0	0

$0\times1 + 0\times2$
$+0\times3 + 3\times4$
$= 12$

12	0¹	0²	0
0	3³	3⁴	0
0	3	3	0
0	0	0	0

$0\times1 + 0\times2$
$+3\times3 + 3\times4$
$= 9 + 12$
$= 21$

12	21	0¹	0²
0	3	3³	3⁴
0	3	3	0
0	0	0	0

$0\times1 + 0\times2$
$+3\times3 + 0\times4$

12	21	9	0
0¹	2³	3	0
0³	4³	3	0
0	0	0	0

$0\times1 + 2\times3 +$
$0\times3 + 4\times3$
$6 + 12$

12	21	9	0
18	3¹	2³	0
0	3³	4³	0
0	0	0	0

$3\times1 + 2\times3$
$+3\times3 + 4\times3$
$3 + 6 + 9 + 12$
$= 30$

12	21	9	0
18	30	3¹	2⁰
0	3	3³	4⁰
0	0	0	0

$3\times1 + 3\times3$
$3 + 9$
$= 12$

12	21	9	0
18	30	12	0
0	1	2	3
0	3	4	0

$3\times2 +$
$6 +$

12	21	9	0
18	30¹	+12	0
6	3⁴	2³	0
0	0³	4⁰	0

$3\times1 + 2\times3$
$3 + 6$
$= 9$

12	21	9	0
18	30	12	0
6	9	3¹	2⁰
0	0	0³	4⁰

$3\times1 + 0$
$= 3$

12	21	9	0
18	30	12	0
6	9	3	0
0	0	0	0

Convolution Eq $x(m)\ h(q-m)$

Length of 1st sequence $= l_1$

" " 2nd " $= l_2$

" " output " $= l_1 + l_2 - 1$

First sequence starts at $n = n1$

Second sequence starts at $n = n2$

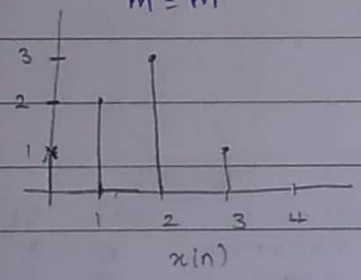
output sequence starts at $n = n_1 + n_2$

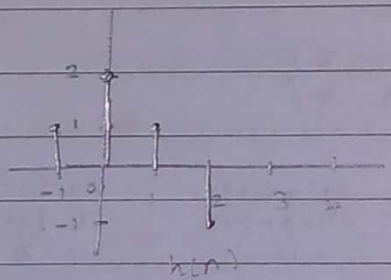
the " " ends at $(n_1 + n_2) + (l_1 + l_2 - 2)$

Q. $x(n) = \{\underset{\uparrow}{1}, 2, 3, 1\}$ (0 1 2 3) $\quad h(n) = \{\underset{\uparrow}{1}, 2, 1, -1\}$ (-1 0 1 2)

starts starts

① $n = m$





$l_1 = 4 \quad l_2 = 4$

result length $= l_1 + l_2 - 1 = 4 + 4 - 1 = 7$

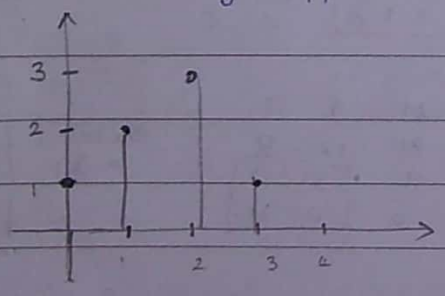
$n1 = 0 \quad n2 = -1 \quad n = n_1 + n_2 = 0 - 1 = -1$

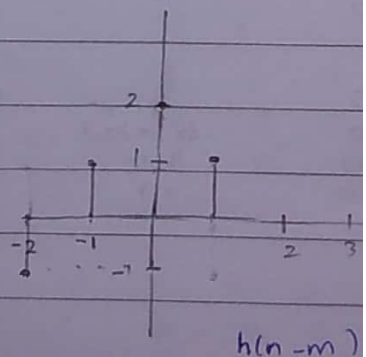
ends at $(n_1 + n_2) + (l_1 + l_2 - 2)$

$(0 - 1) + (4 + 4 - 2) = 5$ ② $x(n) = x(m)$

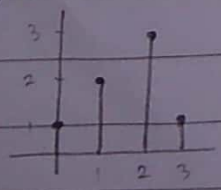
③ $n = -m$ change happens only in h signal

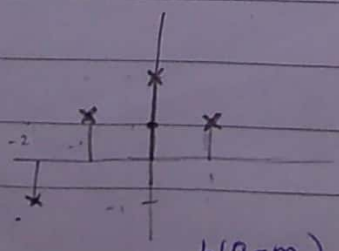
④



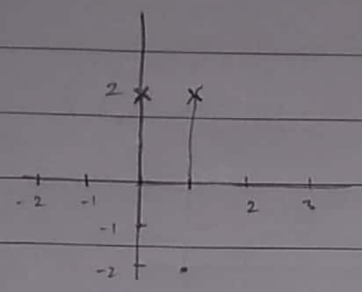


$q = 0$



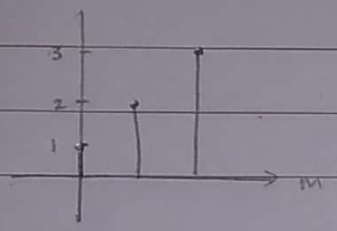


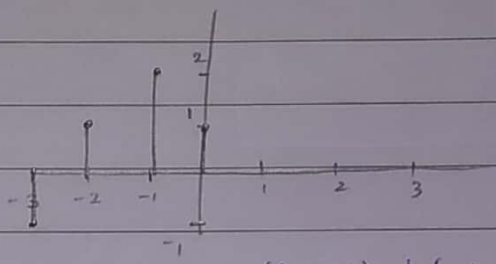
$x(m) \cdot h(0-m)$



4

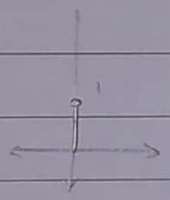
(5) $q = -1$





~~$h(1-m)$~~ $h(-1-m)$

$x(m) \cdot h(-1-m)$



final answer $\{1, 4, 8, 8, 3, -2, -1\}$