Data	visualization (plotting)
back	(plothy)

Page

	Data Vistating)
edi.	
	TARREST TO A WIN
	Is noise related to Xz
	Data Visualization : (1) mus (1) elevis well yes
*	Data Visualization
	la with a visual represon
	of data and is part of data analysis.  of data and is part of data analysis.
	of data and is part of such into a stranslating data into a stranslating data into a stranslating data into a
10 H	chart, graph or other visual components.
	chart, graph or
	a projected ed are 23200 and 1
L. *	variables (featives)
	-> variables refer to characteristics, properties, or
	attributes that can be measured, observed, or
	recorded for a particular entity or unit within a
	dataset.
	Variables -> Dependente xiq elan to #
	independent
	ATIB WEST TEALWREET IEM GIFF
	QUALITATIVE CUANTITATIVE
	(categorical) (Numerical)
٠	> Describes the quality
	Universidate Analysis -> classes
*	Travariant interest of the contract of the con
	conly considering I feature at a time.
2 2 15/20	contains only 1 attribute or a data type that
aller .	contains only 1 attribute or characteristic.
	THURS ON ( .

	Histogram: Frequency Distribution Graph
	Box Plot: Compare the Spread of the variables and
	get an insight into outlier.
	Pared St. A. C.
*	Bivariate Analysis: (Remember: If one var influeences the change
	in the other variable, then you have an independents & dep var.
	-> Mainly used to compare two sets of data
	to find a relationship blu the two variables.
	asseril singer is mi 2408 1-1
	scatter plot, heat map, contour plot, pair plot
	2 dingle House Hit Zimpy potoisi
	- Scatter Plot: Captures the correlation b/w the two
*	Multi-variate Analysis
	-> used to reveal the relationship among several
	variables simultaneously.
12	> Assists in making informed decisions by considerity
	multiple variables & their interactions.
	The leaves of the second of th
	-> Ex: Grouped Box Plot, Multi-variate Scatter Plot, 3D Scatter plot
*	Visualization Techniques
	o surporte a sold to had go
•	Distribution of data points: Box plot, Histogram
	V
•	Comparis on of data points: Multi-line chart, Bar plot,
0	Relations hip
	Correlation of data points: Scatter Plot
•	Composition of datapoints: Pie chart, stacked Area chart/Barchart

	Pair Plot
	Pair Plot  -> preliminary idea  data to find the
	given data to ma the
62 T 3	-> Preliminary idea  -> Prelim
	-> Pair plot visualizes given dates relationship b/w them and plots pairwise relations
F. R. Min. J. C.	the cretationship blewing
	SILTIS USE di Tori
	multiple variables at on cl
187	
	-> Diagonal subplots mare the mivariate
201	(i) D10 C2
	histograms for each attribute.
000	wistograms for each attribute  of f diagonal entries are the scatter plots.
*	Joint Plot 212plant poteins value
Fresh	s Joint plot combines univariate and bivariate
	Plots to visualize relationship b/w 2 variables
Marsh 20	is It consists of a scatter plot for the bivarial
	relationship, with additional marginal plots
	for each variable.
さかき アラナラ か	2 08 7019 400 Ex: Front Continue 3 wheel Plat 30 5
	-> Helps understand correlation & distributions
	of two variables simultaneously.
0	Distribution of data ports: But Place mistration
	10 (501) (10 11) W.
**	14 red traits and there : storing sent to me zive; many
7 701.	The state of the s
	Josephan of data points sales Plate

Page

	1 Lucia France
*	Heatmap
	is color-coded representation of a 2D data,
	representing magnitude of individual values
	within a dataset.
	Contract to the track of the tr
	-> Colors are used to represent the magnitude,
9"	intensity with the color gradient scheme ranging
in the	from a lighter color (low) to darker colors (high)
	Mile or isor south to send to
	as Displays the correlations or relationships in
	a correlation matrixiones Musico 2 cond
*	Parallel co-ordinates
- 2	- Select the most relationsh subsect of teath
	- Parallet co-ordinates allows for the comparison of
	multiple data records, by using parallel lines to
	Connect points based on multiple numerical variables
افتيار	iple with most mediamed information ford the color
	-> Each vertical line is a dimension.
	133 pc 28 = C
	is A data : tem is connected by line segments.
	of no successor with to secon
	Is large number of samples clutters the visualization.
	crodonia - compared to the compared the
	TOTAL TOTAL OF STREET OF S
	10 2 10 10 10 10 10 10 10 10 10 10 10 10 10
	possesses white stable according
	Petal width Sepal Sepal Mico 3 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
	471 0 11 m r 30 - 11 10 20 39 28

Dimensionality
to have for a dataset.
with of i/p variables or features for a dataset;
referred to as its dimensionality.
$y = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_{30} x_{30}$
- Disticultes related to training machine learning
models due to high dimensional data
moders are to dimensionality
=> Curse of dimensionality
Dimensionality Reduction
Dimensionality Reduction
1) Feature selection
-> select the most relevant subset of features
-> Reducing the number of irrelevant features.
or promount of property of 22 most paper by the
@ feature extraction
-> Extracking / deriving information from the origine
features set to create a new features
Subspace.
-> compress data with the goal of maintaining
most of the relevant info.
head to see the seemple adjusted the seemple -
Forward-Feature selection
Features: 1 2 3 4 10
10
Each feative cat
Each feature set we get an avg some.
mare with highest as
teabor 1
· Repeat till termination. check arg scort

-1	
1	pridación de la company de
1	FORWARD FEATURE
1	-> It iteratively selects one feature at a time,
1	evaluating the model's performance after adding
1	each feature & keeping/removing the best subset of
	features that maximizes/minimizes the chosen
1	performance metrics.
	10 Nondedon K
	5,10
	5,8,10,4 20119 devolution to 20 149 24 242
	5, 7, 8, 10
	3000 (ct = 500000 ) modes (ct = 500000
	-> computationally cheaper as it starts with no feabres,
	and only a few features are needed to reach optimal
-	performance.
-	Box PLOT Tolucks mediam 2 was am
	is works well when number of features is very high,
-	as it starts small & adds only informative features.
-	
	-> Since features are added one by one, its easier to
- Annual Property of the Parket of the Parke	track the contribution of each new feature.
	-> only adds " the "best" feature at each step without
	considering combination of features that might work
	well together later to the man interest in the desire
	=> miss the best overall set of features.
	To their are mary features to evaluable F. S can

be slow, especially if the model is complex.



## Backward Feature Selection

- removes one feature at a time, and evaluate the model's performance
- -> If the perf improves, we keep the feature removed; Otherwise, we add it back.

in a product to the south of strong

The final set of features that maximizes or minimizes
the chosen performance metric is returned as the
selected feature subset.

PCA: Sinear transformed me &co. X. B. M. E. Syl Smith

12901×3450708 9400 2222 Larre port 10

×1345678910 200001000

1345678910

Longe harrin and insulations of sure of a gome 21.

- feature interactions that forward selection might miss, thereby leading to a more optimal set of features compared to F-S.
- backward elimination can quickly remove irrelevant ones and reach a good solution.
- -> computationally expensive esp with large datasets or high dimensional feature spaces, as starting with all features means the model has to be fit with a large set of features initially.

NOTICE TO THE PARTY OF THE PART	
	-> starting with all features
New Control of the Co	
	(if there are many irrelevant / redundant feating
	volume to the contract of the
*	Feature Extraction
	180,00 2 661 9 10 26; m 1- 5
	- Aims to reduce # of features in a dataset
2007	by creating new features from the existing on
J	(1900 les cards original mes)
	feature extraction techniques:
	· PCA: linear transformation techniques by finding
	orthogonal axes that capture the most
	Variance.
	0.781281
	· 12 om ap, t-SNE3: Non-linear discording
70 22	technique that remain
	Smy Smy Chro dell 4
2280	The state of the s
	t-distributed stochastic Neighbour Formalli
	stochastic Neighbour Embedding.
	Larenz 21 22 mas 3 12 m
JJ 17 10	Language property and the state of the same of the
	Word-population and the
	For the same of th
23.22	
ettiess	1 consider 25 June 2 Ju
the st	The said of second from the contract to the second to the
- 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1	Market Ma