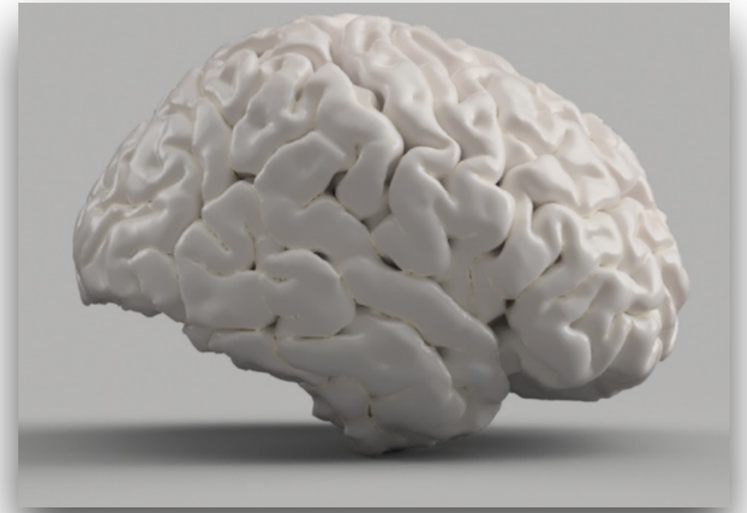
COGS109: Lecture 17



Overview of ML and Modeling approaches

July 27, 2023

Modeling and Data Analysis

Summer Session 1, 2023

C. Alex Simpkins Jr., Ph.D.

RDPRobotics LLC | Dept. of CogSci, UCSD

Plan for the day

- Announcements
- Project updates
- Big picture of machine learning, AI and model structures
- Big picture of modeling
- Dimensionality reduction - brief overview
- Modeling and Data analysis review
- Final thoughts

Announcements

- Reminder on student evals - if >85% do a review class gets a 1% bonus
- We have released feedback on most assignments - please check and if there are regrade requests please politely reach out to Sagarika and cc me
- Assignments remaining D5, D6, D7-EC, Q4, project
- Project meetings to check in

Modeling types and classes

How AI fits into it all

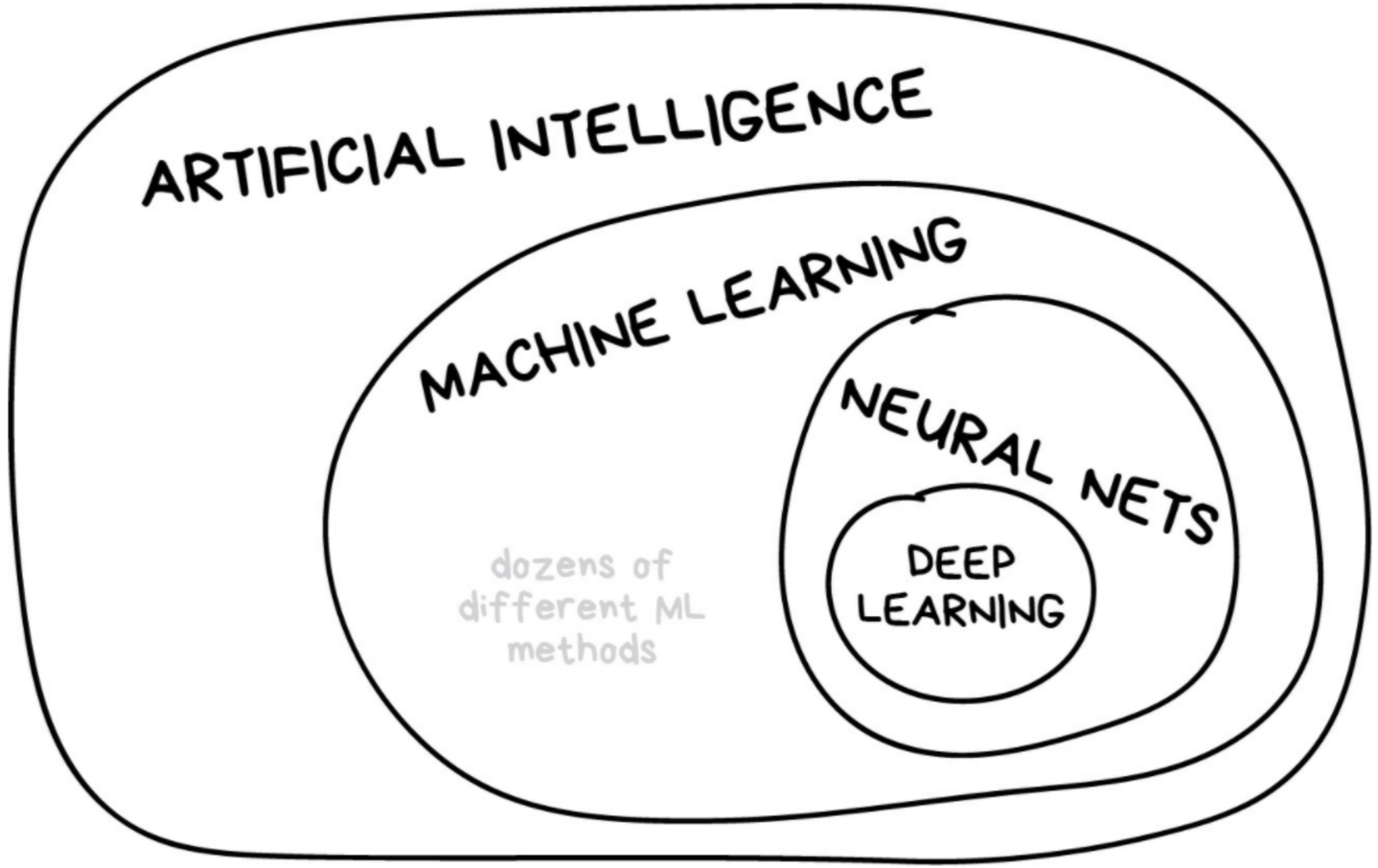


Image source: https://vas3k.com/blog/machine_learning/

The main types of machine learning

- **Classical ML** - simple data, known characteristics/features
- **Reinforcement learning**- no data but we can provide the teaching signal (usually environmental testing and interaction)
- **ANNs** - complicated data, unknown black box, don't care how decisions are made
  - **Deep learning** - solve the problem by breaking it down into simpler sub-parts that are partial solutions
- **Ensembles** - averaging over several solutions to solve a particularly noisy/difficult problem, reduce variance
  - Can also be done with neural networks, but not always

CLASSICAL MACHINE LEARNING

Data is pre-categorized or numerical

Data is not labeled in any way

SUPERVISED

Predict a category

Predict a number

CLASSIFICATION

«Divide the socks by color»

REGRESSION

«Divide the ties by length»

UNSUPERVISED

Divide by similarity

Find hidden dependencies

Identify sequences

CLUSTERING

«Split up similar clothing into stacks»

ASSOCIATION

«Find what clothes I often wear together»

DIMENSION REDUCTION (generalization)

«Make the best outfits from the given clothes»

Image source: https://vas3k.com/blog/machine_learning/

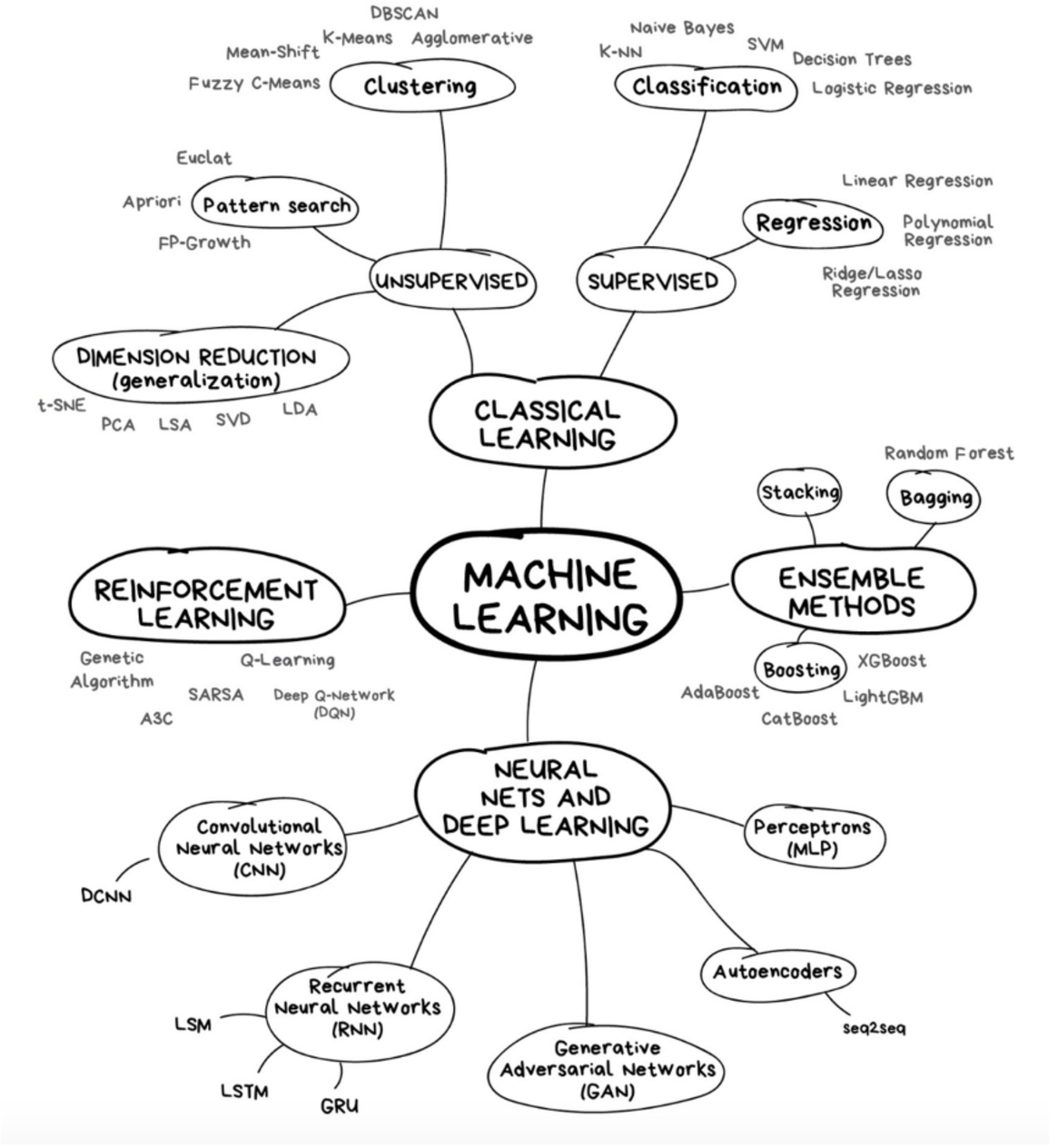
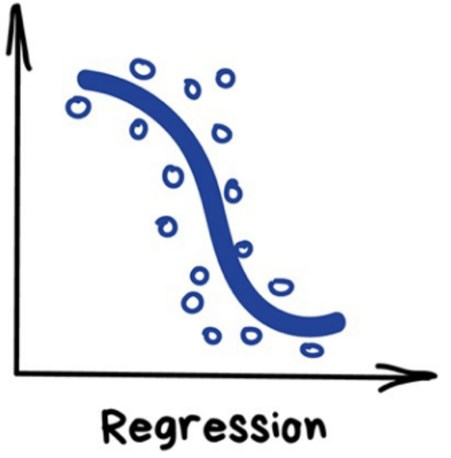


Image source: https://vas3k.com/blog/machine_learning/

Regression

"Draw a line through these dots. Yep, that's the machine learning"



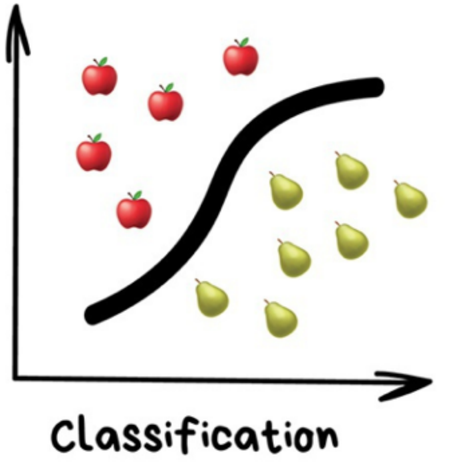
Today this is used for:

- Stock price forecasts
- Demand and sales volume analysis
- Medical diagnosis
- Any number-time correlations

Popular algorithms are Linear and Polynomial regressions.

Classification

"Splits objects based at one of the attributes known beforehand. Separate socks by based on color, documents based on language, music by genre"



Today used for:
– Spam filtering
– Language detection
– A search of similar documents
– Sentiment analysis
– Recognition of handwritten characters and numbers
– Fraud detection

Popular algorithms: Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine

Image source: https://vas3k.com/blog/machine_learning/

When to use Regression vs. Classification

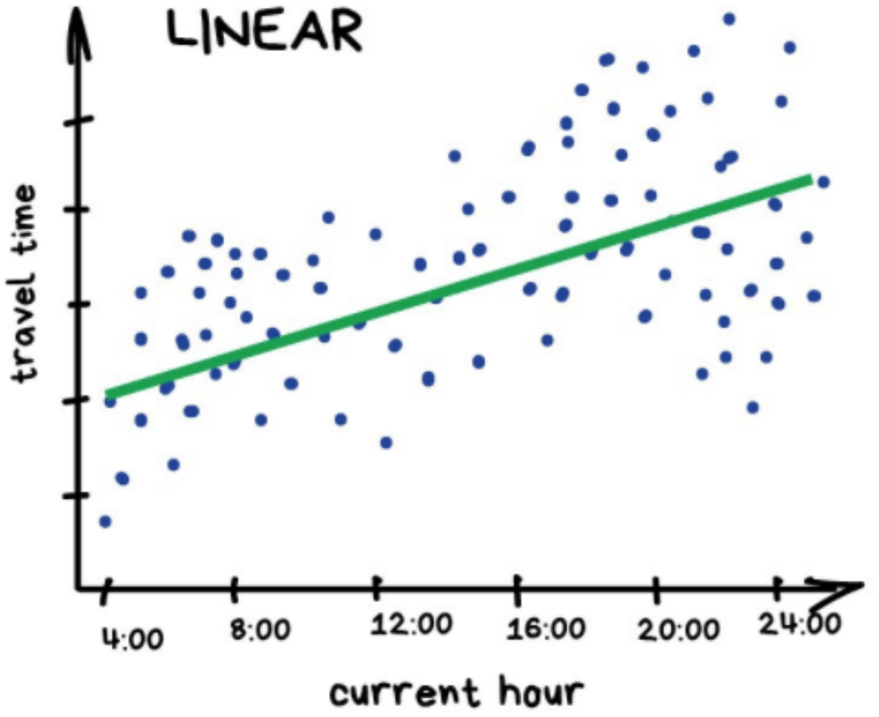
Regression:
predicting continuous variables
(i.e. Age)

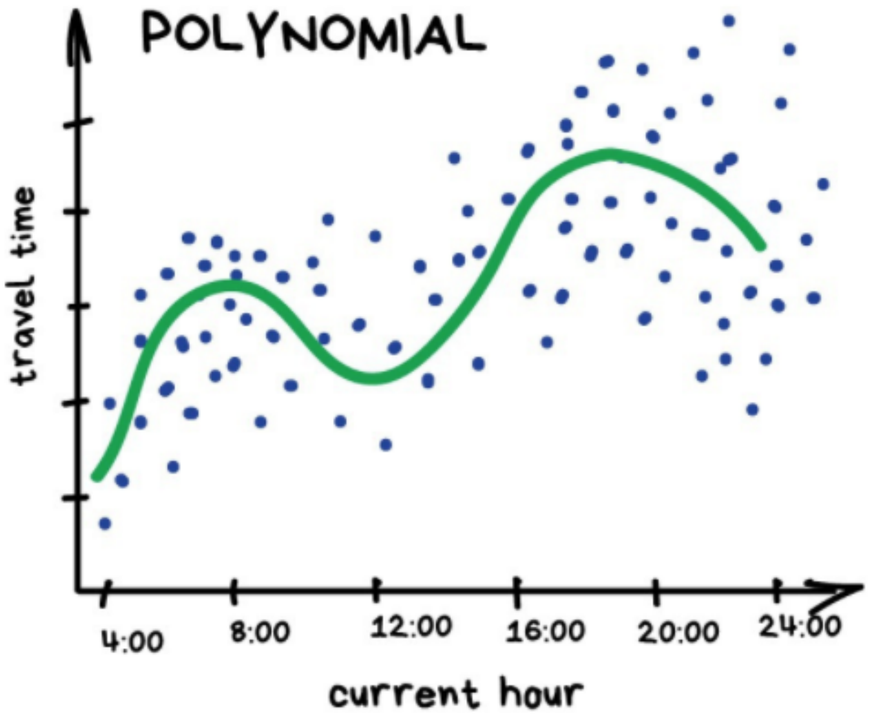
continuous variable prediction

Classification:
predicting categorical variables
(i.e. education level)

categorical variable prediction

PREDICT TRAFFIC JAMS





REGRESSION

Image source: https://vas3k.com/blog/machine_learning/

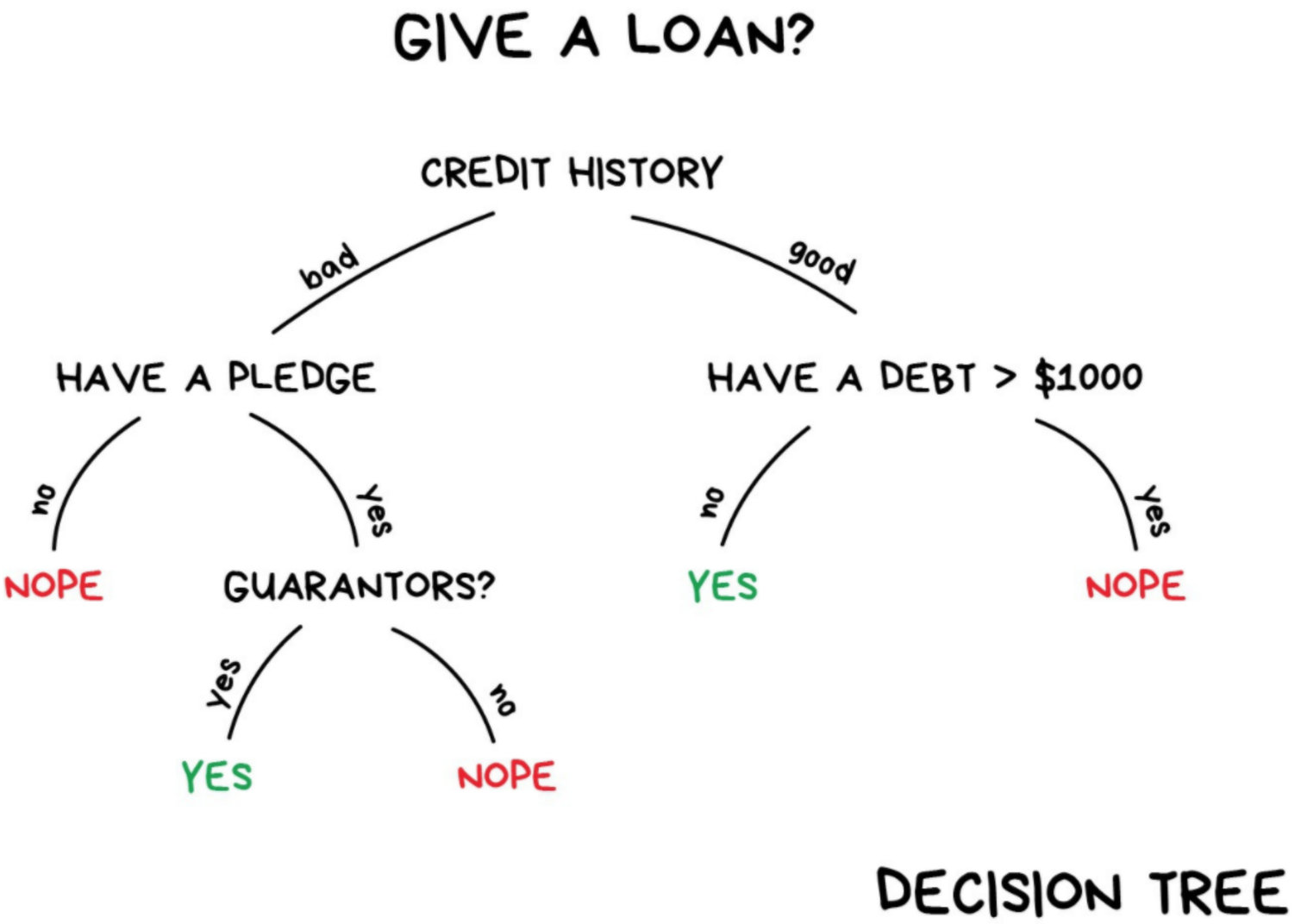
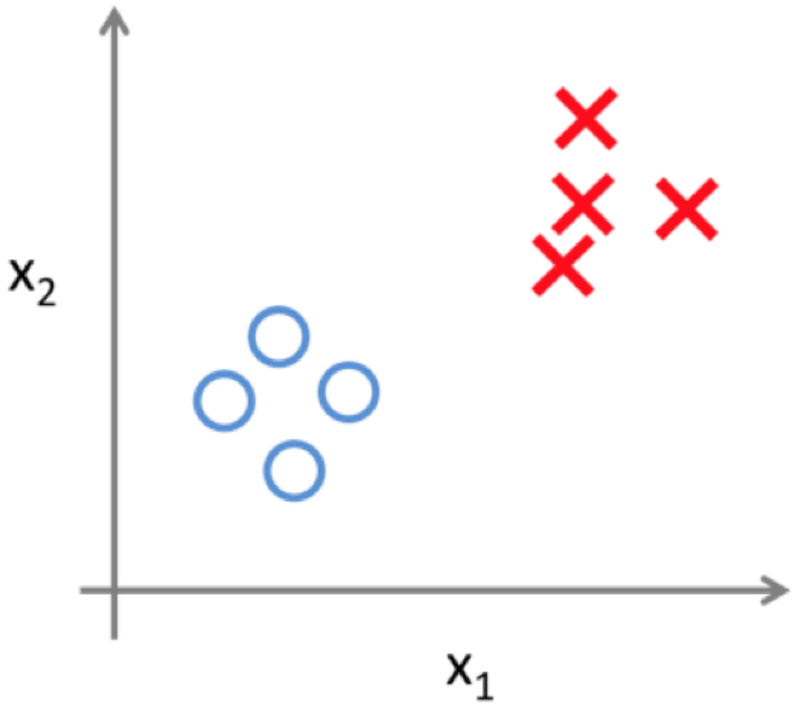


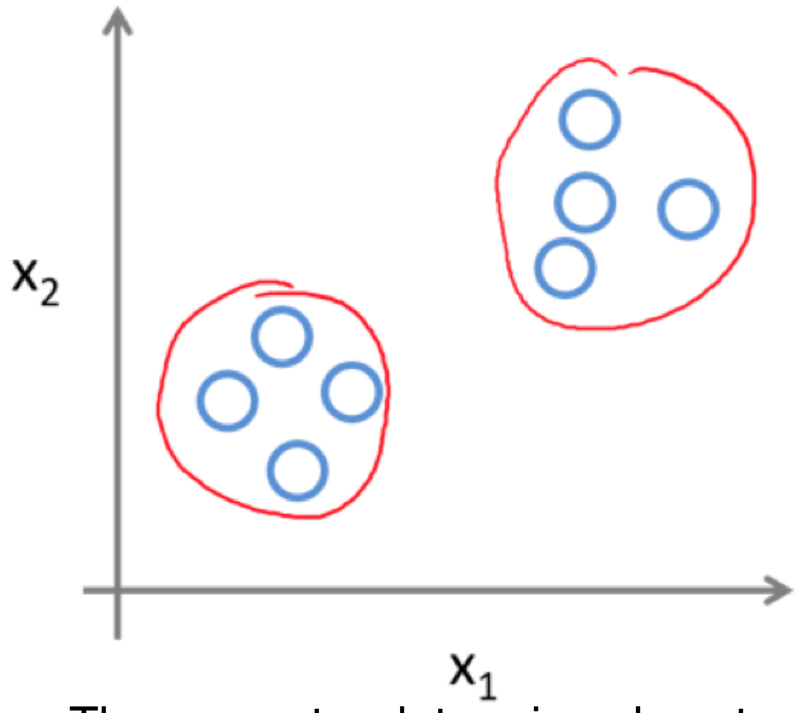
Image source: https://vas3k.com/blog/machine_learning/

Two modes of machine learning

Supervised Learning



Unsupervised Learning



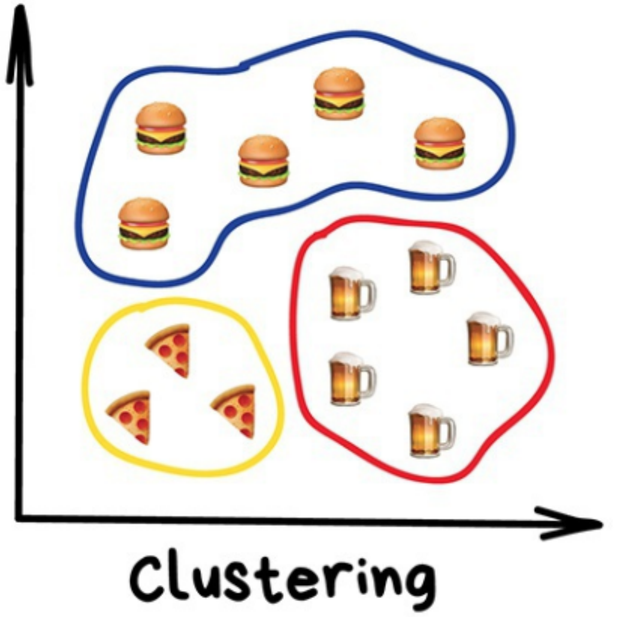
The computer determines how to classify based on properties within the data

Clustering

"Divides objects based on unknown features. Machine chooses the best way"

Nowadays used:

- For market segmentation (types of customers, loyalty)
- To merge close points on a map
- For image compression
- To analyze and label new data
- To detect abnormal behavior

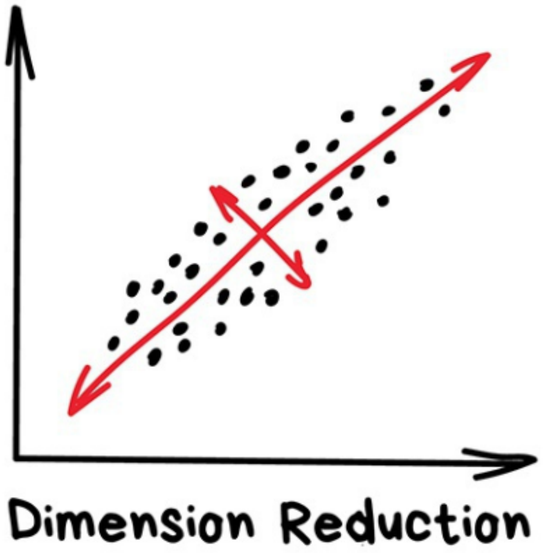


Popular algorithms: K-means_clustering, Mean-Shift, DBSCAN

Image source: https://vas3k.com/blog/machine_learning/

Dimensionality Reduction (Generalization)

"Assembles specific features into more high-level ones"



Nowadays is used for:

- Recommender systems (★)
- Beautiful visualizations
- Topic modeling and similar document search
- Fake image analysis
- Risk management

Popular algorithms: Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Latent Dirichlet allocation (LDA), Latent Semantic Analysis (LSA, pLSA, GLSA), t-SNE (for visualization)

Image source: https://vas3k.com/blog/machine_learning/

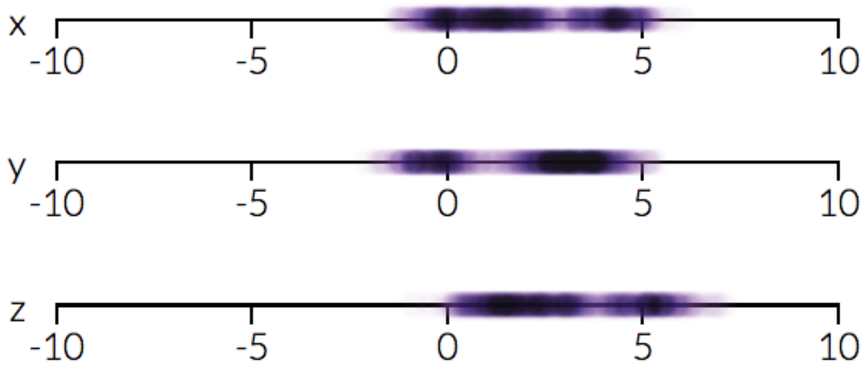
Dimensionality Reduction

A mathematical process to reduce the number of random variables to consider

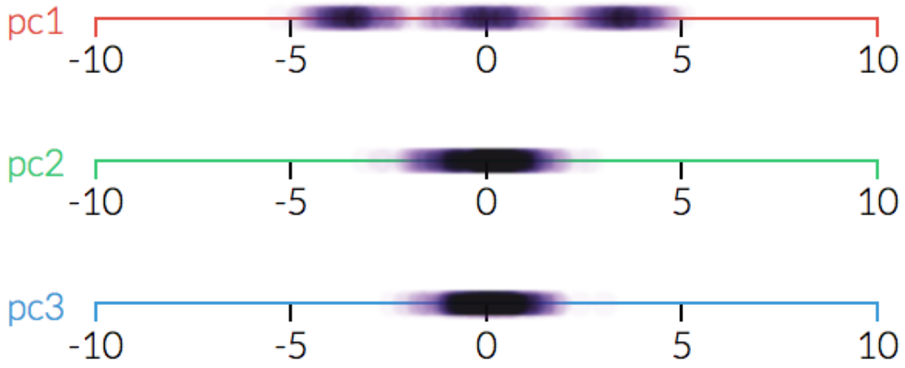
Discuss: why may we want to do this?

Dimensionality Reduction

- Reduce the dimension of quantitative data to a more manageable set of variables



- Reduced set can then be input to reveal underlying patterns in the data and/or as inputs in a model (regression, classification, etc.)



http://setosa.io/ev/principal-component-analysis/

Dimensionality Reduction - Synergies

- How do you control redundant degrees of freedom in a useful way?
    - ***Synergies*** - *coordinated movements that couple a system's degrees of freedom together to reduce control complexity*
- **Importance** - human body has massive redundancy for a given task, and given its compliance the entire body must be actuated to perform simple movements

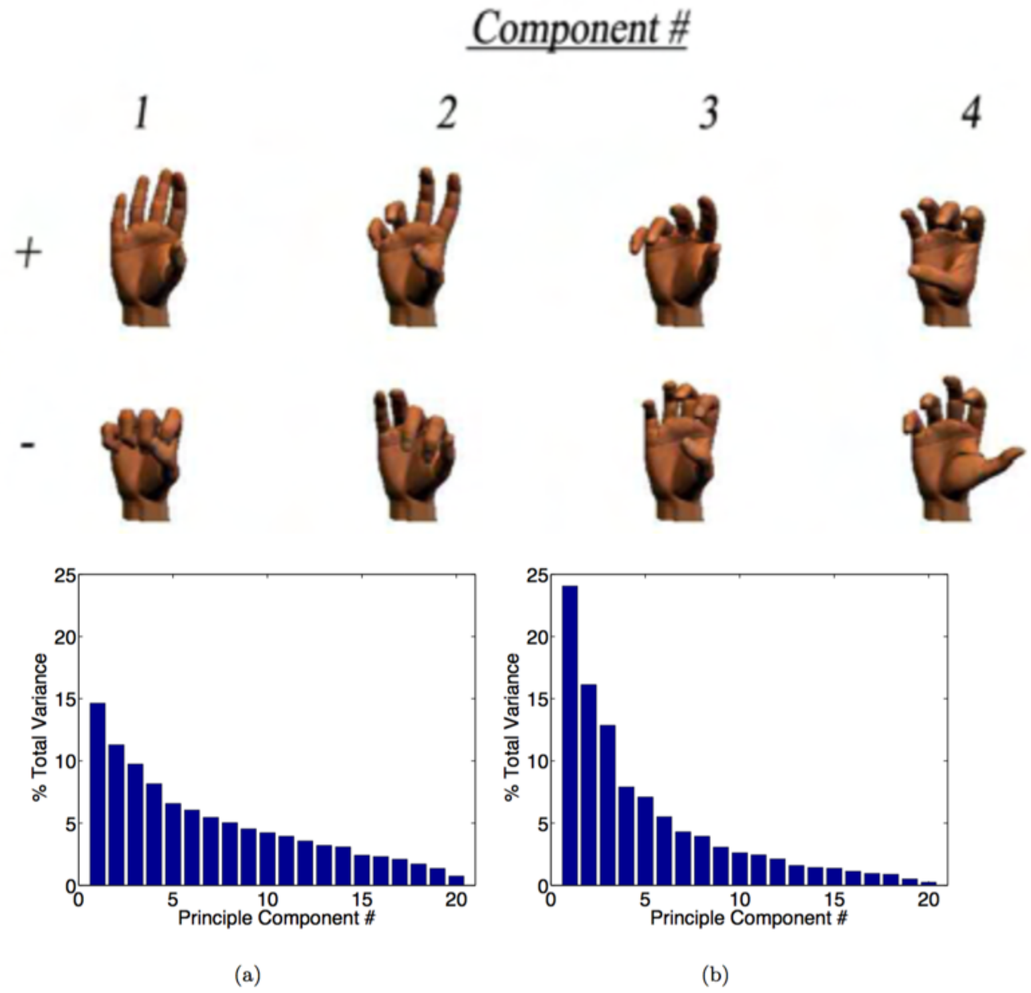


Figure 5.8: Eigenspectrum of the principal components for motor noise and postures. (a) PCA of Covariance of motor noise. (b) PCA of postures Note that the motor noise has a much more flat spectrum than the postures. This is the spectrum for the normalized 0-1 analysis. Other analysis results are similar and thus are omitted for clarity.

Image source: http://casimpkinsjr.radiantdolphinpress.com/files/SimpkinsThesis.pdf

More Use Cases for Dimensionality Reduction

- Thousands of sensors used to monitor an industrial process
    - Reducing the data from these 1000s of sensors to a few features, we can then build an interpretable model
    - Goal : predict process failure from sensors
- Understanding systems of biological redundancy
    - Human movement is in a lower dimensional space relative to our degrees of freedom

Methods for dimensionality reduction

- Projecting high-D data into a lower-D space
- **Methods**
- PCA - Principal component analysis
- ICA - Independent component analysis
- CCA - Canonical correlation analyiss
- Clustering
- FA - Factor Analysis
- (And others!)

Image source: https://www.svds.com/value-exploratory-data-analysis/

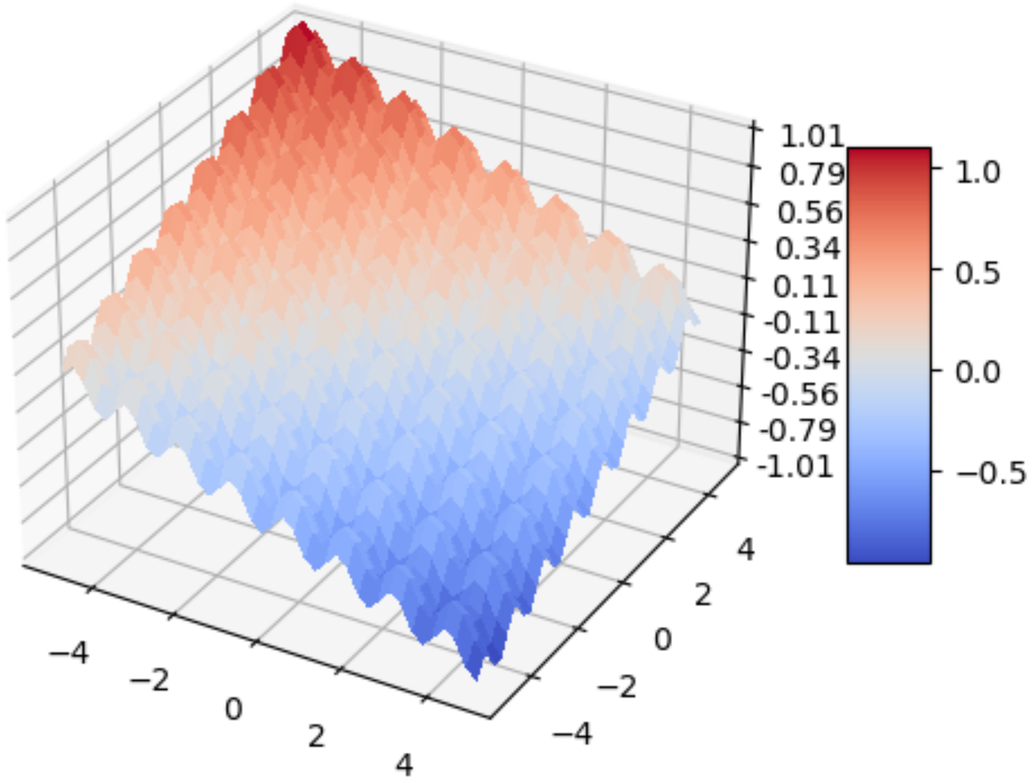
Major example methods

- **PCA** - (Linear) Find projections of the data into lower dimensional space that captures most of the variations in the data
- **ICA** - (Linear) Separate mixed additive independent signals into separate sources
- **CCA** - (Linear) Looks for relationships between two multivariate data sets
- **Clustering** - (Nonlinear) We have discussed this - uses machine learning to extract features from data

Image source: https://www.svds.com/value-exploratory-data-analysis/

So big picture

- Multivariate data usually occupies a lower dimensional subspace, or a slice that captures most of the features of the data
- The question is how do we find that slice?
- Typically some sort of multidimensional rotation



- Note: generated with wk4_dimreduction.ipynb

Principal Component Analysis (PCA)

Key Terms:

- **Principal Component (PC)** - a linear combination of the predictor variables
- **Loadings** - the weights that transform the predictors into components (aka weights)
- **Screeplot** - variances of each component plotted

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

Principal Component Analysis (PCA)

Goal : combine multiple numeric predictor variables into a smaller set of variables. Each variable in this smaller set is a weighted linear combination of the original set.

This smaller set of variables -- the ***principal components*** **(PCs)** - “explain” most of the variability of the full set of variables….but uses many fewer dimensions to do so.

The **weights (loadings)** used to form the PCs explain the relative contributions of the original variables to the new PCs.

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

“Simple” PCA : Two predictor variables (X_1 and X_2)

For two variables X_1 and X_2 there are two principal components Z_i with i=1 or 2

$$Z_i = w_{i,1}X_1 + w_{i,2}X_2$$

$w_{i,1}$ and $w_{i,2}$: weightings (*loadings*)

- Transform the original variables into principal components

Z_1 : the first principal component (PC1)

- The linear combination that best explains the total variance

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

S&P 500 Data: 5648 days (1993-2015) x 517 stocks

	ADS	CA	MSFT	RHT	CTSH	CSC	EMC	IBM	XRX	ALTR	ADI	AVGO	BRCM	FSLR	INTC	LLTC	MCHP	MU	NVDA
1/29/93	0	0.06012444	-0.0220998	0	0	0.01889746	0.00736807	0.0921652	0.25914009	-0.0071053	-0.0157849	0	0	0	-0.0504878	-0.0898696	0	0.03702057	0
2/1/93	0	-0.180389	0.02762115	0	0	0.01888884	0.01842489	0.11520651	-0.1007745	0.06389288	-0.0157929	0	0	0	0.09536733	0.0449348	0	0.03702038	0
2/2/93	0	-0.1202566	0.03589987	0	0	-0.0755726	0.02948172	-0.0230413	0.02879553	-0.0141924	0.0473628	0	0	0	0	0.0674022	0	0.12340155	0
2/3/93	0	0.0601242	-0.024857	0	0	-0.151128	0.00368875	-0.2534543	-0.04319	-0.0071053	0.20523612	0	0	0	-0.050495	0.0224674	0	-0.0123403	0
2/4/93	0	-0.3607697	-0.0607567	0	0	0.11335029	-0.0221136	0.0698618	0	-0.0070962	-0.0315699	0	0	0	0	0.0224674	0	-0.0740409	0
2/5/93	0	0.03005777	0.09389247	0	0	0.09445283	-0.0479066	0.04657454	0.17276006	-0.0212976	-0.0631478	0	0	0	-0.0476873	-0.0674022	0	-0.0123403	0
2/8/93	0	0.03006643	-0.0607498	0	0	-0.1133503	-0.0110568	0.11643635	-0.04319	0.00709618	0	0	0	0	-0.0196321	-0.1235743	0	-0.0617008	0
2/9/93	0	-0.0901902	-0.063521	0	0	-0.1322391	-0.0147456	0.06986181	-0.115169	0.04969143	-0.0157929	0	0	0	-0.0112235	0.0224674	0	0	0
2/10/93	0	0.12025657	0.02209981	0	0	0.09445283	0.01474557	-0.2561599	0.01439448	0.02838473	0.01578495	0	0	0	0.04487956	0.11233699	0	0.07404095	0
2/11/93	0	0.03005825	-0.0220927	0	0	-0.0188975	0.01474556	-0.1397236	-0.04319	0.02129762	-0.0315699	0	0	0	-0.0532953	0.06740222	0	-0.0246804	0
2/12/93	0	-0.0901901	-0.0358999	0	0	-0.0377863	-0.0073681	-0.0698618	-0.1871546	0	-0.0473628	0	0	0	-0.0336561	-0.112337	0	-0.0370204	0
2/16/93	0	-0.6313411	-0.0607497	0	0	-0.0377863	-0.0479066	-0.0931491	-0.04319	-0.0283938	-0.1262955	0	0	0	-0.098175	-0.1460417	0	-0.0246803	0
2/17/93	0	0.12025657	-0.0165712	0	0	-0.1700254	-0.0110568	0.04657453	-0.08638	-0.0142015	0.03157785	0	0	0	0.04487955	0	0	-0.0123403	0
2/18/93	0	-0.1803808	0.00828562	0	0	-0.0566751	0.00368875	-0.0931491	-0.08638	0	-0.0157849	0	0	0	-0.0168315	0.0224674	0	0.03702056	0
2/19/93	0	0.03006595	-0.0469427	0	0	0	0.00736807	-0.0232873	0.115169	0.01419237	0.03157785	0	0	0	0.10378311	0.15727183	0	0.14808196	0
2/22/93	0	0.03005825	-0.0662782	0	0	-0.1322477	-0.0184249	0.13972361	0	0.02839382	-0.0631557	0	0	0	-0.0168317	-0.0674022	0	0	0
2/23/93	0	-0.0300583	0.03314266	0	0	0	-0.0479066	-0.0698618	-0.1439645	-0.0070962	0.03157785	0	0	0	-0.0336631	-0.0337047	0	-0.0493606	0
2/24/93	0	0.15031459	0.10769942	0	0	0.01888884	0.04421782	0.1397236	0.08638003	0.00709618	0	0	0	0	0.11781411	-0.0224674	0	0.09872137	0
2/25/93	0	0.15032277	0.04142827	0	0	0.01888884	-0.0110568	0.37259628	0	0.02839382	0	0	0	0	0.0112163	0.15727183	0	-0.0370205	0
2/26/93	0	-0.0300659	-0.0193286	0	0	-0.0188888	0.01105682	0.06986181	0.05759106	0	0.01578495	0	0	0	-0.028055	-0.0224674	0	-0.074041	0
3/1/93	0	-0.180381	-0.0497068	0	0	-0.0944614	-0.0073681	0	0.04358505	0.00709618	0.09472561	0	0	0	-0.0336631	0.0224674	0	0	0
3/2/93	0	0	0.06351413	0	0	0.15113659	0.00368875	-0.0698618	0.116229	0	0.11051053	0	0	0	0.09537435	-0.0449348	0	0.12340155	0
3/3/93	0	0.12025658	0	0	0	-0.0566751	0.03684977	0.16301088	0.02905891	-0.0070962	0	0	0	0	0.01402383	-0.112337	0	-0.0370204	0
3/4/93	0	-0.1503146	-0.0220927	0	0	0.0377863	0.00367932	-0.0698618	-0.1452879	-0.0070962	-0.015785	0	0	0	-0.0252473	-0.0674022	0	0.01234016	0
3/5/93	0	0.03005825	-0.0165714	0	0	-0.0944614	0.00368875	-0.0232873	0	0.03549001	0	0	0	0	-0.0617041	0.0449348	0	0.02468042	0
3/8/93	0	0.06012444	0.02209275	0	0	0.01888884	-0.025793	0.11643634	0.21792524	0	0.04736279	0	0	0	0.06731932	0.13480441	0	0.09872114	0
3/9/93	0	0.09019015	0.00552151	0	0	0.09446144	0.00736807	0.09314908	-0.0290523	-0.0070962	-0.0157849	0	0	0	0.0112163	0.0898696	0	0	0
3/10/93	0	0.03006595	0.01104991	0	0	0	0.01105681	-0.1862981	0.02905891	-0.0141924	-0.0157849	0	0	0	-0.0196321	-0.0449348	0	0	0
3/11/93	0	-0.0300583	0.02761408	0	0	0.22670058	0	-0.1862982	-0.0581112	0.00709618	0.06314774	0	0	0	-0.0196392	0.01123011	0	0	0
3/12/93	0	0	0.06627822	0	0	-0.0188975	0.01474556	0.30273448	-0.1452813	0.02839381	0.06314774	0	0	0	0.02524749	0.01123729	0	0.13574153	0

For this example: we'll focus on 16 top companies

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

Screeplot

The vernacular definition of “scree” is an accumulation of loose stones or rocky debris lying on a slope or at the base of a hill or cliff.

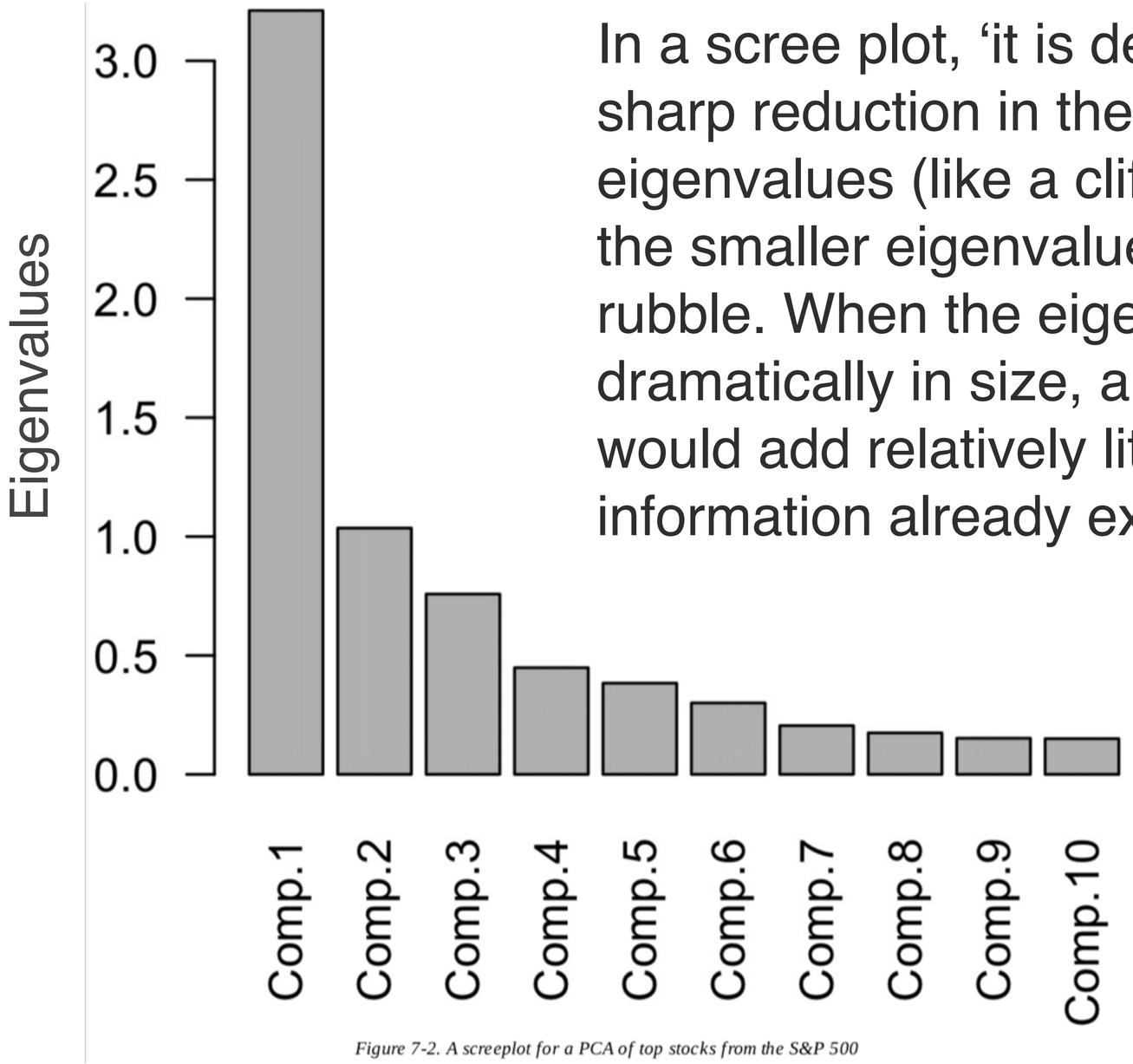


Figure 7-2. A screeplot for a PCA of top stocks from the S&P 500

In a scree plot, ‘it is desirable to find a sharp reduction in the size of the eigenvalues (like a cliff), with the rest of the smaller eigenvalues constituting rubble. When the eigenvalues drop dramatically in size, an additional factor would add relatively little to the information already extracted.’ (Source)

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

Loading of PCs 1-5

PC1: Overall stock market trend

PC2: Price change of energy stocks

PC3: movements of Apple and CostCo.

PC4: movements of Schlumberger to other stocks

PC5: Financial companies

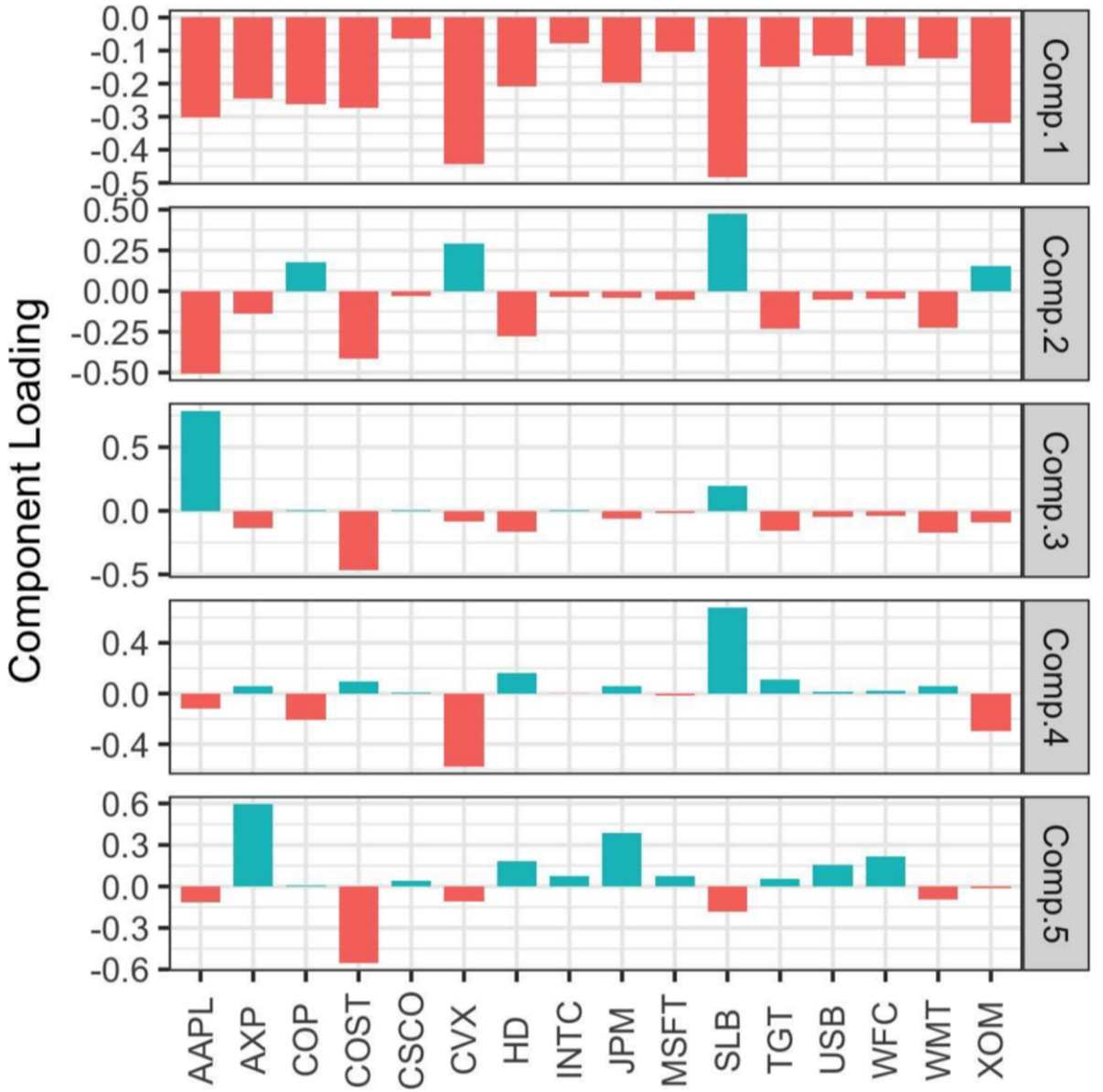


Figure 7-3. The loadings for the top five principal components of stock price returns

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

How many PCs to select?

Option 1: Visually through the screeplot

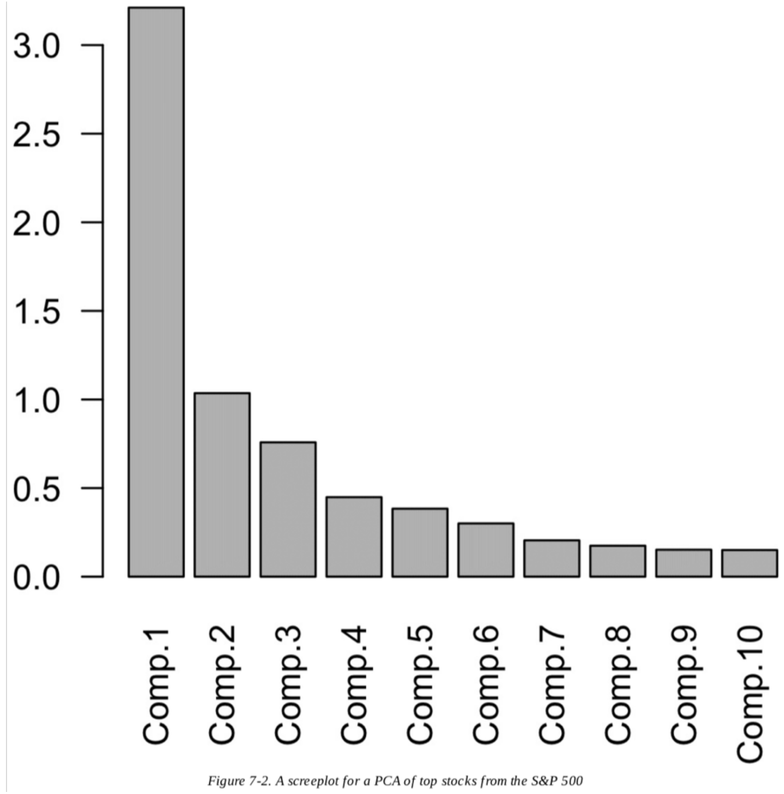


Figure 7-2. A screeplot for a PCA of top stocks from the S&P 500

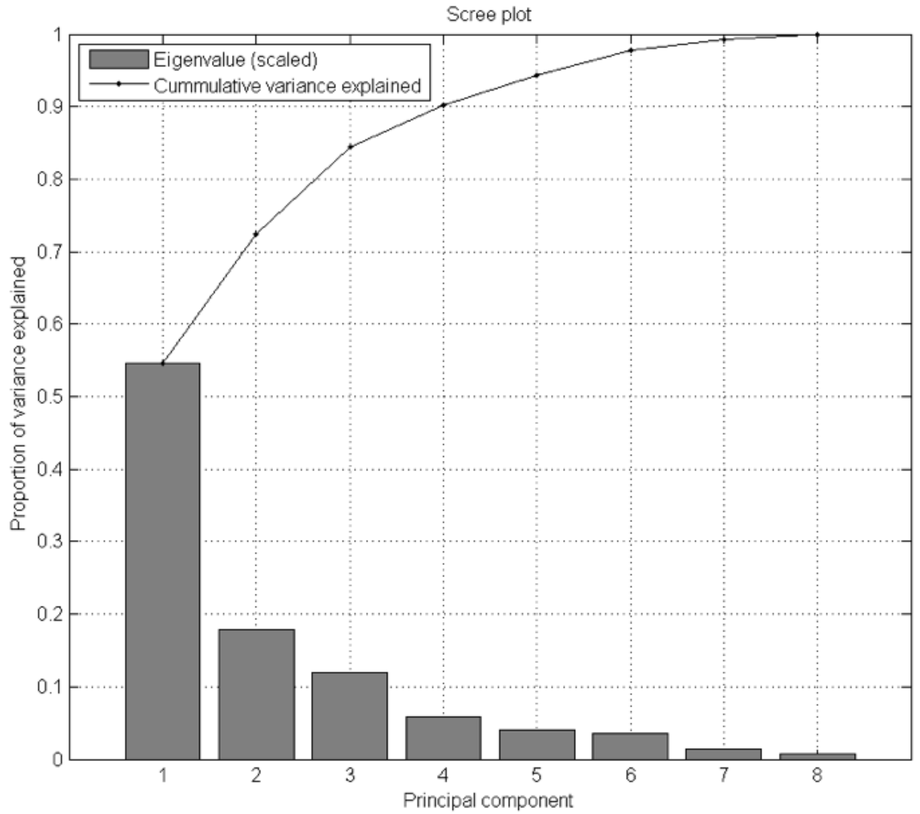
Option 2: % Variance explained (i.e. 80% variance explained)

Option 3: Inspect loadings for an intuitive interpretation

Option 4: Cross-validation

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

Screeplot Interpretations



How many PCs would you likely consider given this screeplot?

A
1

B
2 or 3

C
4 or 5

D
6-7

E
8

PCA : Key Ideas

1. PCs are linear combinations of the predictor variables (numeric data only)
2. Calculated to minimize correlation between components (minimizes redundancy)
3. A limited number of components will typically explain most of the variance in the outcome variable
4. Limited set of PCs can be used in place of original predictors (dimensionality reduction)

For more on PCA:

- https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/
- http://setosa.io/ev/principal-component-analysis/

Source: Practical Statistics for Data Scientists (Bruce & Bruce)

Model structures

- Can be anything represented by a function
  - Equations
  - Heuristics
- Often involve terms you design or are automated
  - Black, Grey, White box modeling
- Use what is the simplest model for your application

Future of modeling

- Discussion