BIG DATA APPROACHES FOR CRITICAL CARE MEDICINE

JOHN R. HOTCHKISS, MD

ASSOCIATE PROFESSOR, UNIVERSITY OF PITTSBURGH SCHOOL OF MEDICINE AND STAFF INTENSIVIST AND NEPHROLOGIST, VA PITTSBURGH HEALTHCARE SYSTEM
PITTSBURGH, PA

John Hotchkiss received his undergraduate and medical degrees from the University of Chicago. He completed medicine residency, nephrology fellowship, and pulmonary research fellowship at the University of Minnesota, as well as a critical care fellowship at Hennepin County Medical Center. He left private practice in Minnesota to join the faculty at the University of Pittsburgh, where he serves as an intensivist, nephrologist, and analyst for the Pittsburgh VA and the Veteran's Engineering Resource Center. Dr. Hotchkiss has had a career long interest in applying quantitative approaches to clinical care in domains ranging from mechanical ventilation to the national allocation of resources. He is board certified in critical care medicine, nephrology, and clinical informatics.

OBJECTIVES:

Participants should be better able to:

1. Be able to list 3 contemporary techniques to extract knowledge from "Big Data;"

2. Be able to outline – in broad terms – how these techniques "work;'

3. Be able to describe – in broad terms – the types of clinical problems for which each of these techniques is most suitable.

THURSDAY, MARCH 23, 2017 9:30 AM

Big Data in Critical Care Medicine

John Hotchkiss, MD
Pittsburgh VA Medical Center

The opportunity

- EMRs, in concert with high density physiologic data streams, are generating enormous volumes of patient data
- These data could
  - Provide valuable insights into disease trajectories
  - Constitute “natural clinical trials” /efficacy illustrations
  - Support individualized precision care

The challenge and the missed opportunity

- Volumes of data are far too large for human processing; automation required
- Administrators and administrative endeavors have been the greatest beneficiaries of the EMR thus far
- Much of the current (recent) effort has been largely "arithmetic" in nature
  - How many widgets are being seen with widget disease X?
  - Are we capturing all of the widgets we see?
  - Are we reporting them appropriately?
  - Are we billing for them appropriately?
- Such work does little to leverage the EMR to benefit patients or society

Goal: outline useful approaches to data analysis that extend <u>beyond arithmetic</u>

- Clustering approaches
- Artificial Neural Networks
- Natural Language Processing

Clustering

In clustering approaches, a distance metric is objectively applied to data in order to

- Elucidate the number of different groups in a set of data
- Classify population members by group
- Serve as a classification tool for individuals not included in the initial data
- Identify features that have high relevance to group assignation

Basic elements of clustering- I

- A dataset of interest comprising labelled members, each of which is cast as an ordered set regarding the germane elements:

Patient1=(65 year old, BMI 41, HbA1c 12, 60 cig/week, mSBP 170, eGFR 35, PO4 4, **slope -4,…**)

Patient2=(34 year old, BMI 20, HbA1c 5, 0 cig/week, mSBP 110, eGFR 55,PO4 4, Slope**-0.15,…**)

Etc

- A metric that can be used to define the distance between any two members of the data set, based on the values of components in their feature vector

Basic elements of clustering- II

- A metric defining the distance between any two members of the data set, based on the values of components in their feature vector:

Euclidean

d(1,2)= [(Age1-Age2)^2 + (BMI1-BMI2)^2 +(A1c1-A1c)^2 + (Cig1-Cig2)^2 +(BP1-BP)^2 + (eGFR1-eGFR2)^2 + +(PO41-PO42)^2 + (Slope1-Slope2)^2] ^(1/2)

Minkowski

d(1,2)= [(Age1-Age2)^**z** + (BMI1-BMI2)^**z** +(A1c1-A1c)^**z** + (Cig1-Cig2)^**z** +(BP1-BP)^**z** + (eGFR1-eGFR2)^**z**+ (PO41-PO42)^**z** + (Slope1-Slope2)^**z**] ^(1/**z**)

Etc.

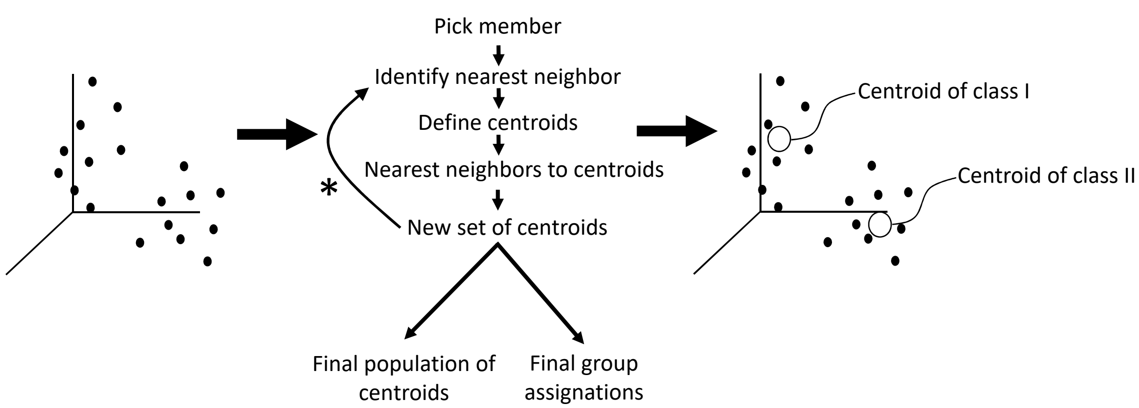
- Distances are used to identify groups and assign individuals to a given group

Two broad approaches to clustering are employed

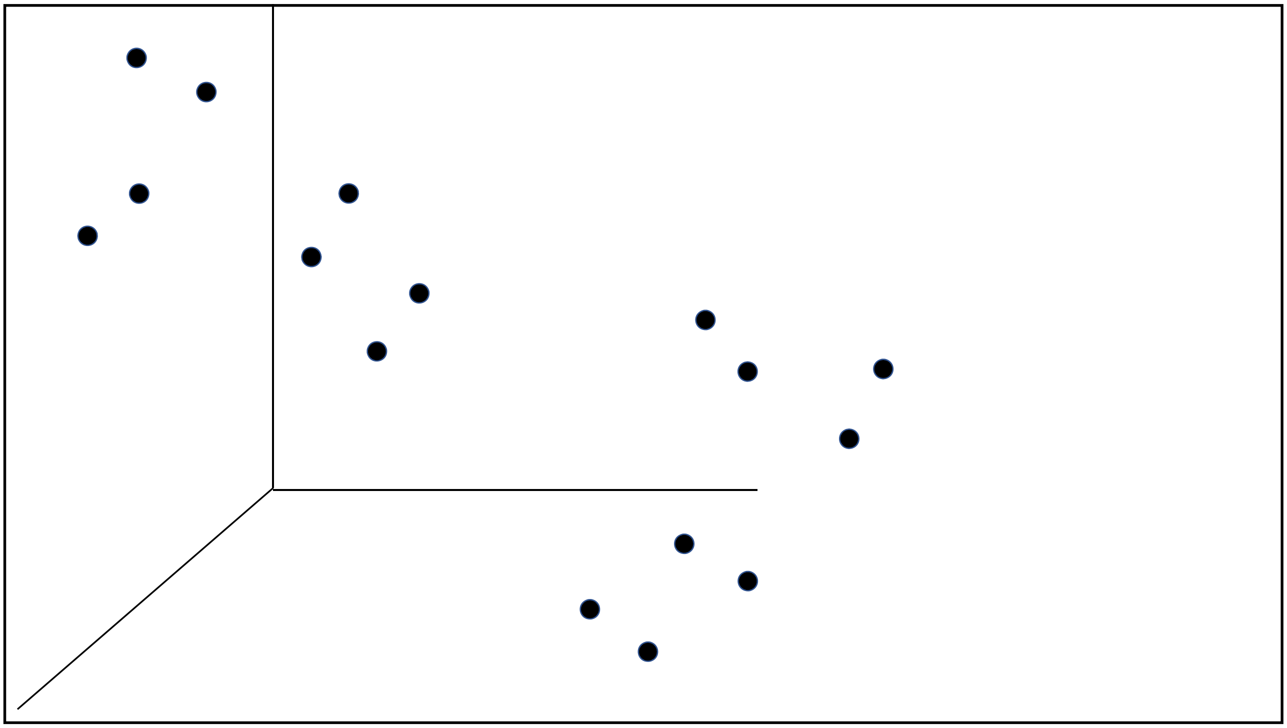
Supervised clustering
- The number of different groups is specified
- The distance metric is applied to partition to overall population into groups
- Example: improved, no change, worsened
- Predicated on knowing the number of groups

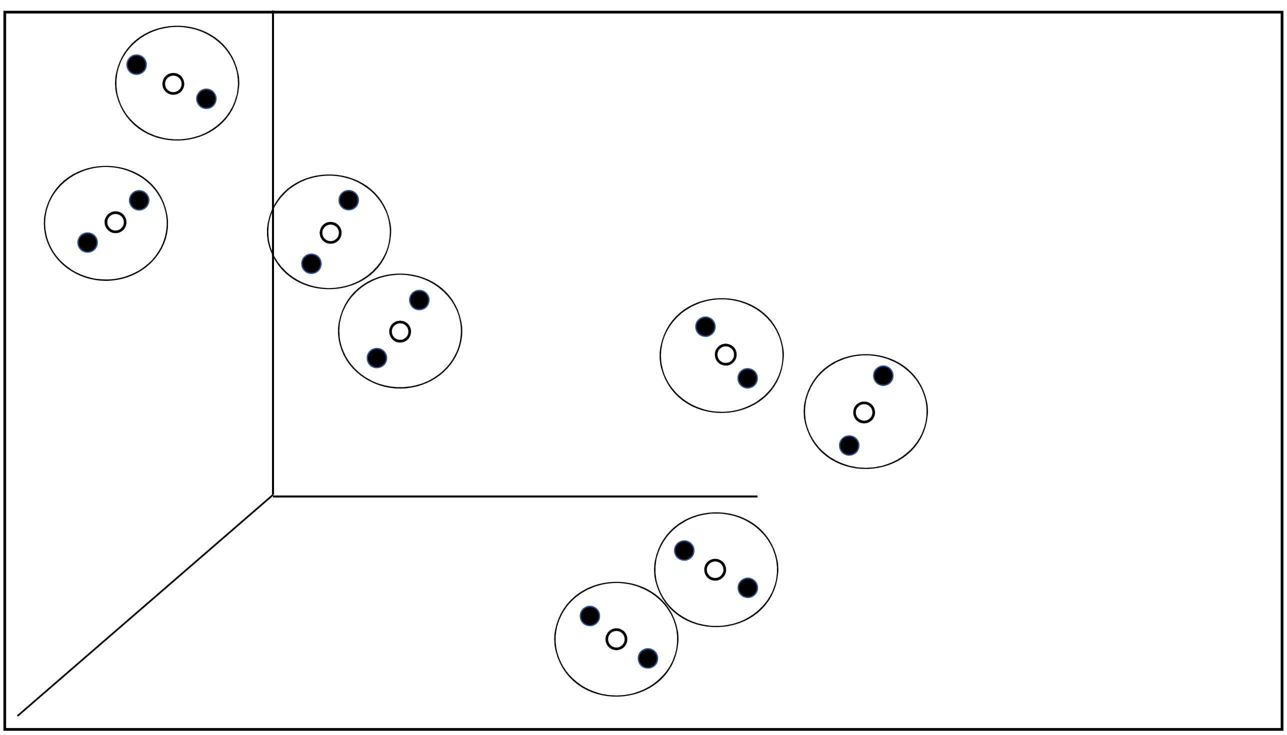
Unsupervised clustering
- The number of groups is not specified
- The distance metric is applied to identify number of groups AND assign individuals
- Useful if the outcome classes are not known
- Less subject to bias

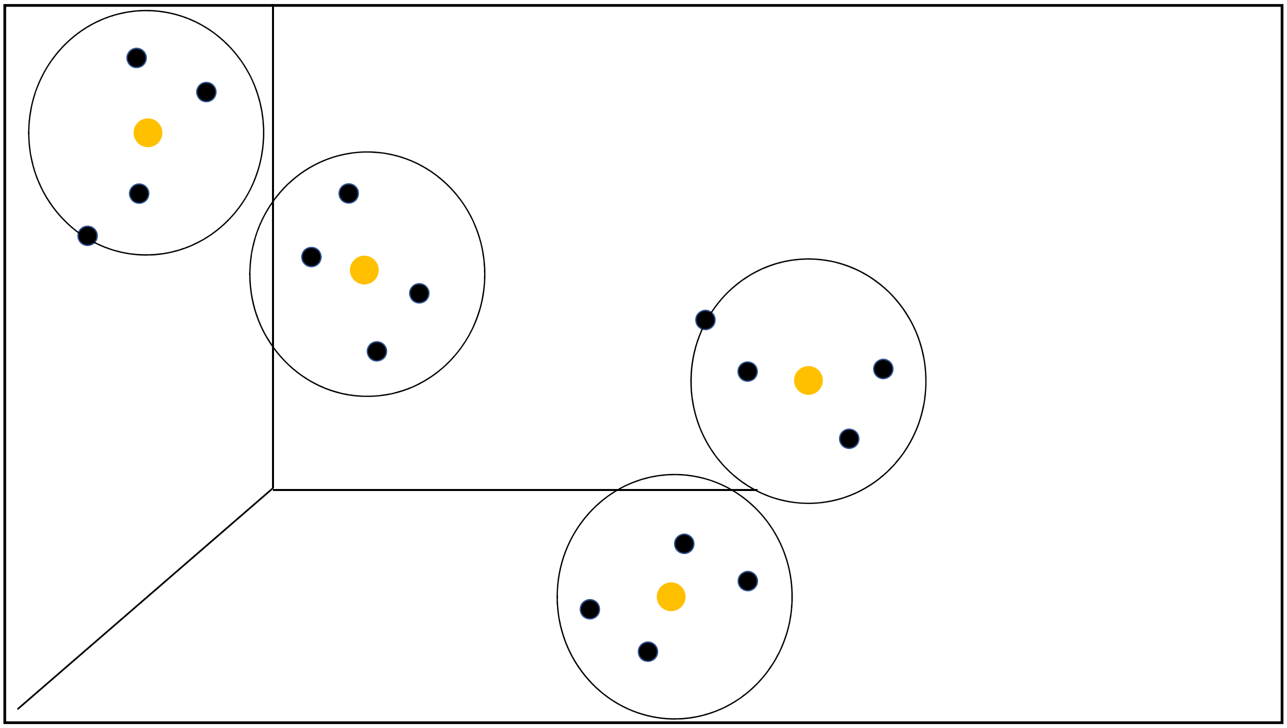
A broad, agglomerative approach based in R^3

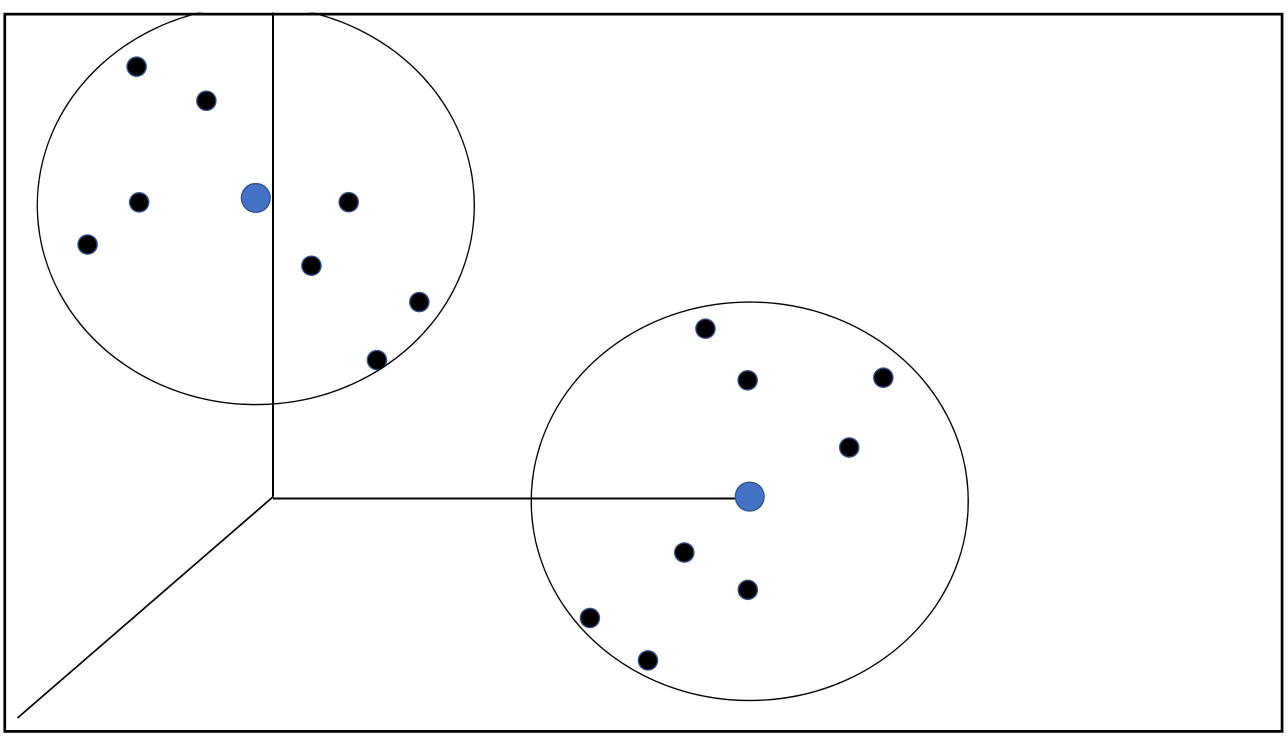


- This is an iterative process (exemplified by *)
- Point of termination is defined by a pre-defined stopping rule
- Example presented is that of agglomerative clustering, rather than a dissection approach



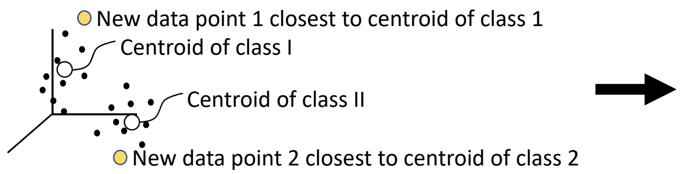






Applications following centroid identification

- In unsupervised clustering, elucidation of number of “different” groups
- In both supervised and unsupervised clustering, use as classification tools to assign individuals to groups
- If groups are associated with particular outcomes, this can be used to assign individuals to risk classes



Individual 1 would be classified as belonging to Group I

Individual 2 would be classified as belonging to Group II

A few loose ends

- Newer methods of clustering exist that do not require exclusive class membership- “individuals” can be partly Class I and partly Class II. This is “fuzzy clustering,” and may be uniquely suited to RWP.
- The clustering approach itself may be employed in an interative fashion:
  - Unsupervised clustering to identify groups
  - Supervised clustering to identify those features most important to effective classification
  - Elimination of superfluous factors and re-clustering to develop optimal classifiers

Practical examples of clustering applications

- Unsupervised: Do patients with CKD "break" into identifiable clusters when administrative data, comorbidities, and relevant laboratory data are reviewed?
  - Yes, there are ~ 6 distinct clusters identified by unsupervised protocol
  - Costs and healthcare resource utilization segregate with specific clusters
- Supervised: Given a known number of observable phenotypes, and a set of phenotype annotated exemplars, identify elements of the feature vectors that optimally assign expected phenotype.

Artificial neural networks

Artificial neural networks are tools used for pattern recognition/ classification and prediction

- Modeled on construct similar to human neural systems, with “Hebbsian” learning => they are trained
- Very trendy in 80s- early 2000s, but seemed that hype overestimated potential & somewhat fell out of vogue
- This may have been unfortunate, as there were limitations in
  - Computing power
  - Training algorithms
  - Network architectures
- With time, these limitations have been addressed and ANN are proving to be extremely powerful tools

Basic elements required for an ANN analysis

- At least one dataset containing a population of exemplars
- An exemplar is a vector of features associated with the pertinent outcome for that vector (individual):

Template:

[Patient, (Outcome: xxx, {Age, BMI, cigs per week, A1c, SBP, eGFR, CRP, albumin})]

[Patient 1, (Died, {Age 54, BMI 43, cigs 50, A1c 6, SBP 170, eGFR 25, CRP 6, albumin 2.8})]

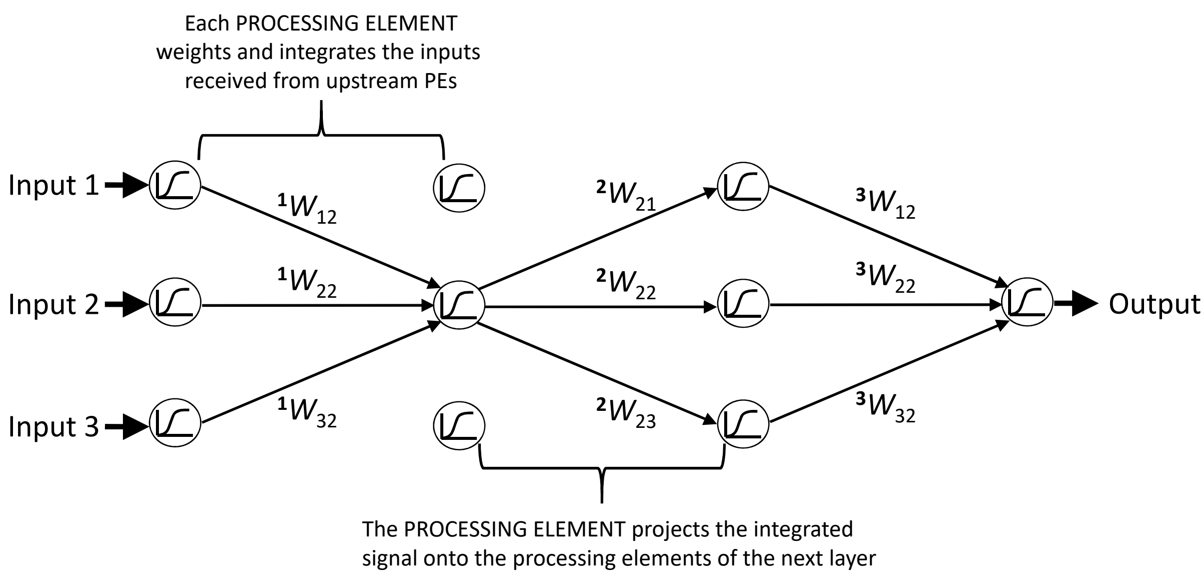
[Patient 2, (Lived, {Age 30, BMI 22, cigs 0, A1c 6, SBP 110, eGFR 60, CRP 1, albumin 4})]

- A rule defining what is to be used as the network’s error metric
- A rule defining when training should be halted

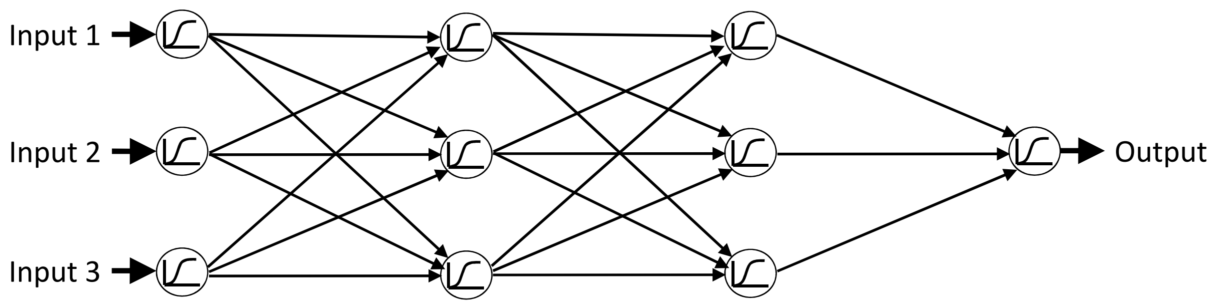
Conceptual elements of an ANN

- Input nodes
  - Composed of processing elements
  - Take as inputs elements of the vector describing the exemplar (e.g., BMI)
  - Connect to the hidden layers
- Hidden layers
  - Composed of processing elements
  - Take as inputs the output of the preceding layer
- Output layer
  - Composed of processing elements
  - Take as input the output from the preceding layer
  - Can function either as a regression tool or as a pure classifier
- Activation function
  - The rule describing how each processing element "handles" its inputs
  - Multiple different types

A picture is worth a thousand words…



Filling in the missing connections



This is a fully connected, non-recursive ANN
with multiple (2) hidden layers

So…what is an activation function?

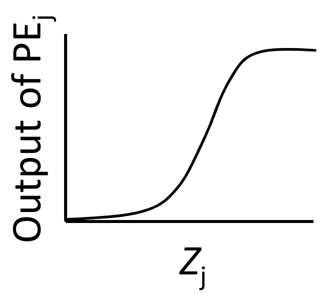
Each processing element forms the weighted sum, Z_j ,of the outputs of the preceding layer and passes this sum through a non-linear transformation, the <u>activation function</u>

The output of the processing Element is passed on to the next network layer

Weighted sum

$$Z_j = \sum_i w_{ij} \times zi$$

Activation function



How are ANN trained (how do they “learn”)?

- The form of the activation function is generally uniform and selected from a common “library” of such non-linear transformations
- Accordingly, classical training is based on the selection of appropriate values for each of the weights connecting successive processing elements
- This involves using a set of exemplars, each with a known associated outcome
- An error function is specified, as is a stopping criteria

This is a recursive process

Training exemplars:
Exemplar 1: [Outcome $_1$, {feature vector $_1$}]
Exemplar 2: [Outcome $_2$, {feature vector $_2$}]
Exemplar 3: [Outcome $_3$, {feature vector $_3$}]
etc

For each exemplar,
input feature vector to ANN
compute error (ANN output vs known outcome)

Calculate total error as function of known outcome and ANN output

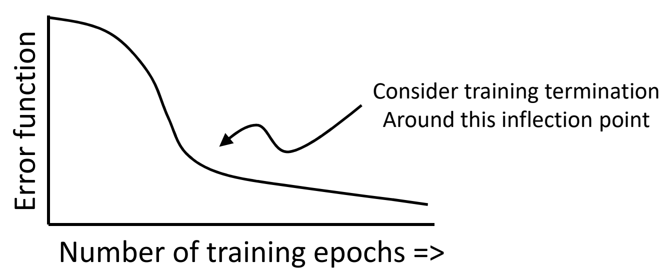
For each connection weight, compute change in total error
As a function (derivative) of the change in connection weight

Revise connection weights and repeat process

Continue loop until stopping criteria is reached

Comments on error function and stopping criteria

- If the mean squared error (MSE) is used and the output is scaled to [0,1], the ANN provides the posterior probability conditioned on feature vector
- Stopping criteria are generally based on a plateau in the error function with successive iterations of reweighting
  - Stop to early => poor performance on the data set at hand
  - Stop too late => excellent performance on DSAH, but overfitting and poor generalization
- In general, for a fully connected ANN, one will need 2 – 10 exemplars per connection weight



So…why bother with ANN?

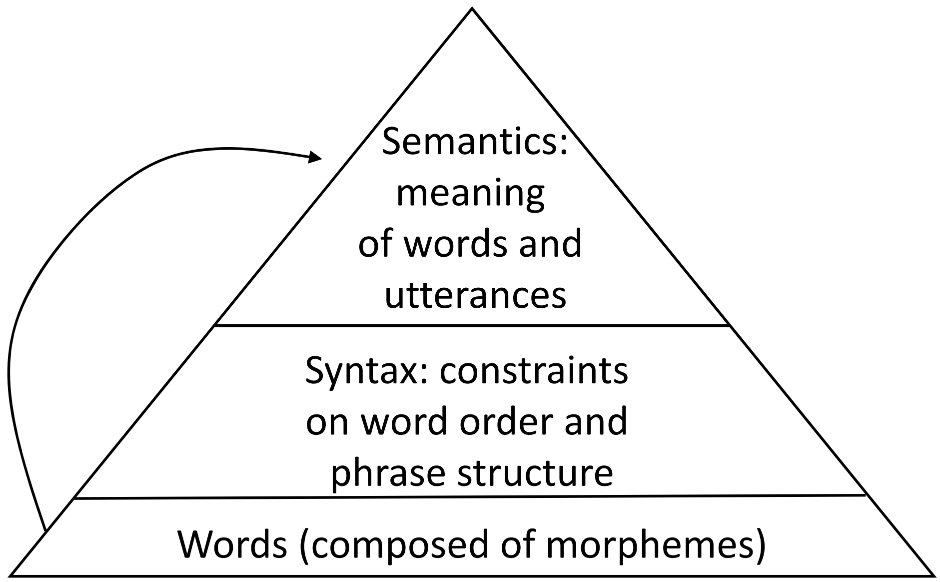
- Very powerful "general mapping" tools, as a multilayer ANN can emulate any definable multivariate function, thus partition or decision surface (Arnold)
- If the number of features (inputs) exceeds 6-8, an ANN will outperform conventional regression analyses (Cybenko)
- With more nuanced architectures (partially connected or convolutional), ANN have demonstrated dramatic capabilities in pattern recognition
- With newer approaches to training, ANN can serve as general learning machines

Natural language processing

Natural language processing

- Differs from clustering and artificial neural networks in that it (generally) addresses the use of terms or words, rather than numerical data
- Much information- in medical records and elsewhere- is in unstructured (uncoded) textual form which is not amenable to automated processing
- NLP can be thought of as rendering such textual or narrative records “computable,” or amenable to automated processing and analysis

A few levels of natural language processing



A basic application of NLP: information extraction

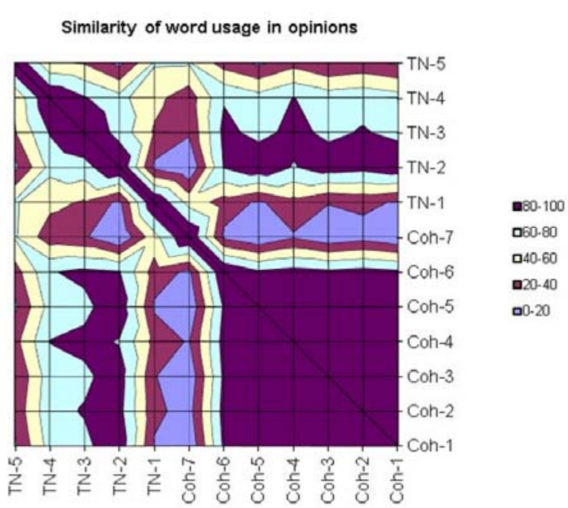
- Based on identifying patterns in text
- Mechanism of pattern recognition
  - key words
  - definition/identification of concepts
- Pattern occurrences can be contextualized and quantified
- Examples
  - Biosurveillance: queries/mentions regarding influenza to identify hotspots
  - Epidemiologic applications: identifying aspects of CKD care in text record
  - Pharmacovigilance: identifying adverse drug events in text record

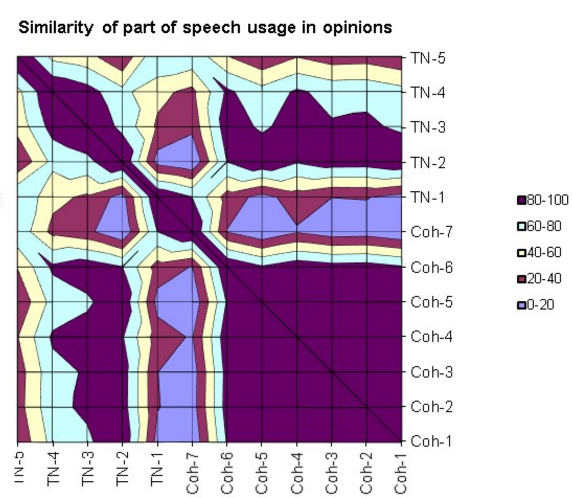
Simplest extraction of information from text: bag of words (or, slightly more complex, bag of phrases)

- For example, seek text string "obstructive sleep apnea" in medical record
- If found, tag record as containing this phrase
- If found, count as one instance [need rules for counting- per record, per instance, etc]
- Subject to syntactic complications: "Does **not** have obstructive sleep apnea" versus "obstructive sleep apnea"
- Often requires establishing a dictionary of essential synonyms (e.g., "OSA.")

BOW can be used to compare document similarity

Example: comparison of legal opinions written by different justices





More sophisticated approaches make use of structures (~ syntax) as well as words/morphemes

- Returning to the BOW example, consider
  - Patient has **obstructive sleep apnea**
  - Patient does not have **obstructive sleep apnea**
  - Based on polysomnography, we suspect the patient has **obstructive sleep apnea**
  - Polysomnography does not support a diagnosis of **obstructive sleep apnea**
- The context in which a word (or phrase) is found can substantively change the meaning of the word or phrase
- This is the key problem addressed by more sophisticated approaches to NLP

Major forms of (more advanced) NLP

- Those based on formal knowledge of linguistics
  - Exemplified by – for example – the work of Chomsky
  - Use linguistic structures generally based on parts of speech:
  - [Determiner|Adjective(s)|Singular noun(s)|Singular or pleural noun(s)]
  - Very much rules based
- Statistical NLP
  - Not based on rigorous application of regular structures
  - Instead, evaluate the common patterns occurring in a corpus
  - "You shall know a word by by the company it keeps" [Firth, 1957]
- Hybrids

More nuanced approaches to NLP => more powerful applications

- Identification of document class and separation (as in applying OCR to scanned documents => character streams => identification of components of individual records)
- Annotation or summarization of documents
- Identification of interrelationships between entities in documents
- Sentiment or opinion identification from text record

The last 40 minutes in a nutshell, with no math

- Clustering approaches
  - Identify natural groupings in a universe of individuals with diverse characteristics
  - Can used to assign individuals to groups, identify important characteristics, or both
- Artificial neural networks
  - Can be applied as either pattern recognition or stratification tools
  - May have more of a “predictive” flavor to them
  - (not discussed) may incorporate feedback
- Natural language processing
  - Renders unstructured textual data “computable”
  - Accordingly, many applications in modern medicine

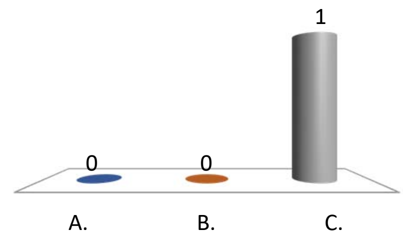
So…Some questions

You are interested in elucidating how frequently PCPs address a (known)diagnosis of asthma in their progress notes. Which approach might prove most useful?

A. Clustering
B. Artificial neural networks
C. Natural language processing

You are interested in elucidating how frequently PCPs address a (known)diagnosis of asthma in their progress notes. Which approach might prove most useful?

A. Clustering
B. Artificial neural networks
C. Natural language processing

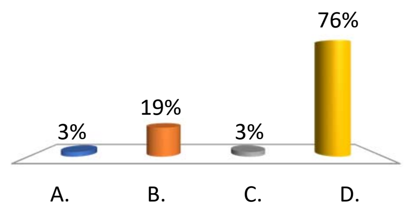


You wish to develop a model predicting likelihood of death during intensive care, based on admission data. Which approach might prove most useful?

A. Natural language processing
B. Artificial neural networks
C. Clustering
D. Either B or C, depending on the desired output

You wish to develop a model predicting likelihood of death during intensive care, based on admission data. Which approach might prove most useful?

A. Natural language processing
B. Artificial neural network
C. Clustering
D. Either B or C, depending on the desired output

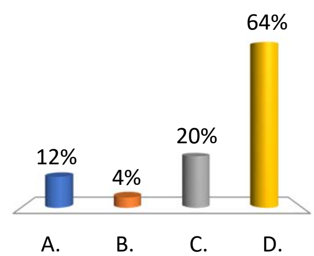


You want to determine patient characteristics that are associated with OSA treatment compliance. Which approach might be most appropriate?

A. Artificial neural networks
B. Natural language processing
C. Clustering
D. Either A or C

You want to determine patient characteristics that are associated with OSA treatment compliance. Which approach might be most appropriate?

A. Artificial neural networks
B. Natural language processing
C. Clustering
D. Either A or C



Thank you for your attention