NATURAL LANGUAGE PROCESSING IN CLINICAL MEDICINE

Frank Rudzicz

UHN Toronto Rehabilitation Institute

University of Toronto

TODAY

- I'm going to tell you:
  - How to organize your data,
  - How to extract features from those data,
  - How to use those features in machine learning, and
  - How talking about cookies can reveal dementia.

ALZHEIMER'S DISEASE

- Le *et al.* (2011) looked for signs of AD in 3 British novelists.

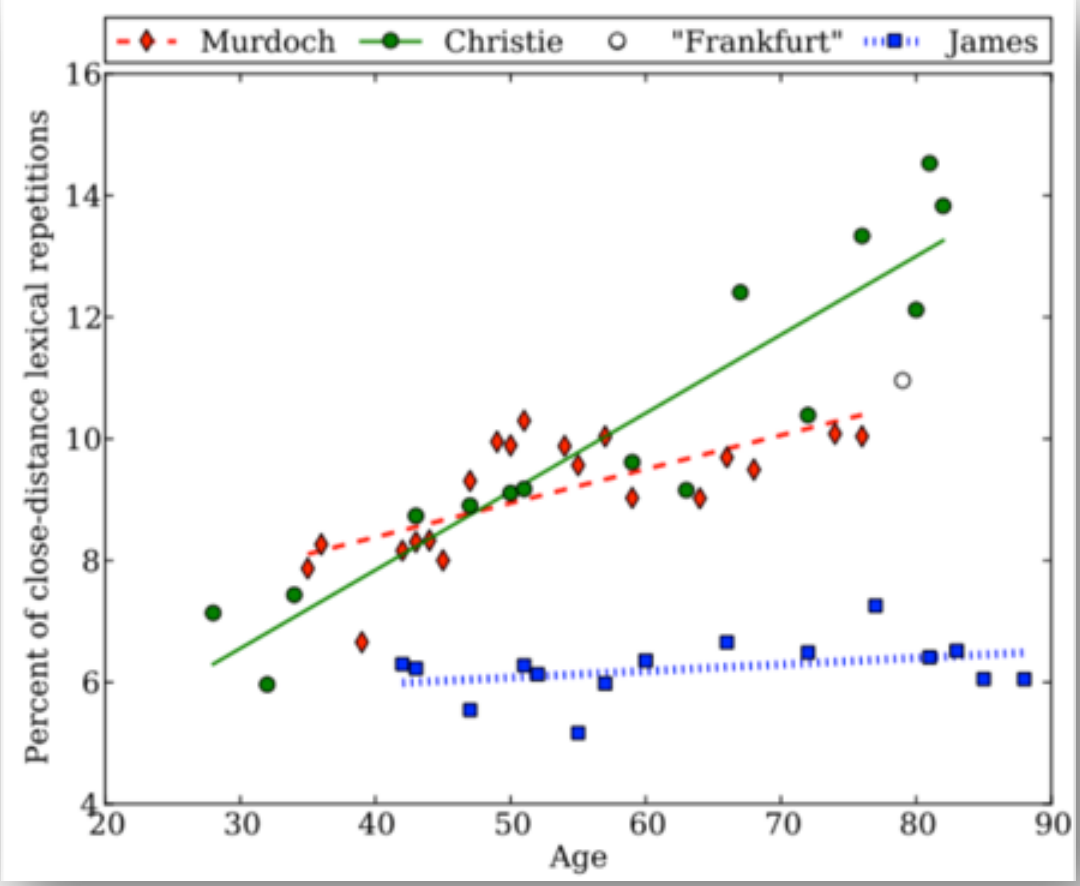
Iris Murdoch
20 novels, ages 35-76
Died of AD

Agatha Christie
16 novels, ages 28-72
Suspected AD

P.D. James
15 novels, ages 42-82
No AD

- Software computed various linguistic measures, e.g.,
  - *vocabulary size, lexical repetition, syntactic complexity, passive voice,…*

FEATURES OF ALZHEIMER'S DISEASE

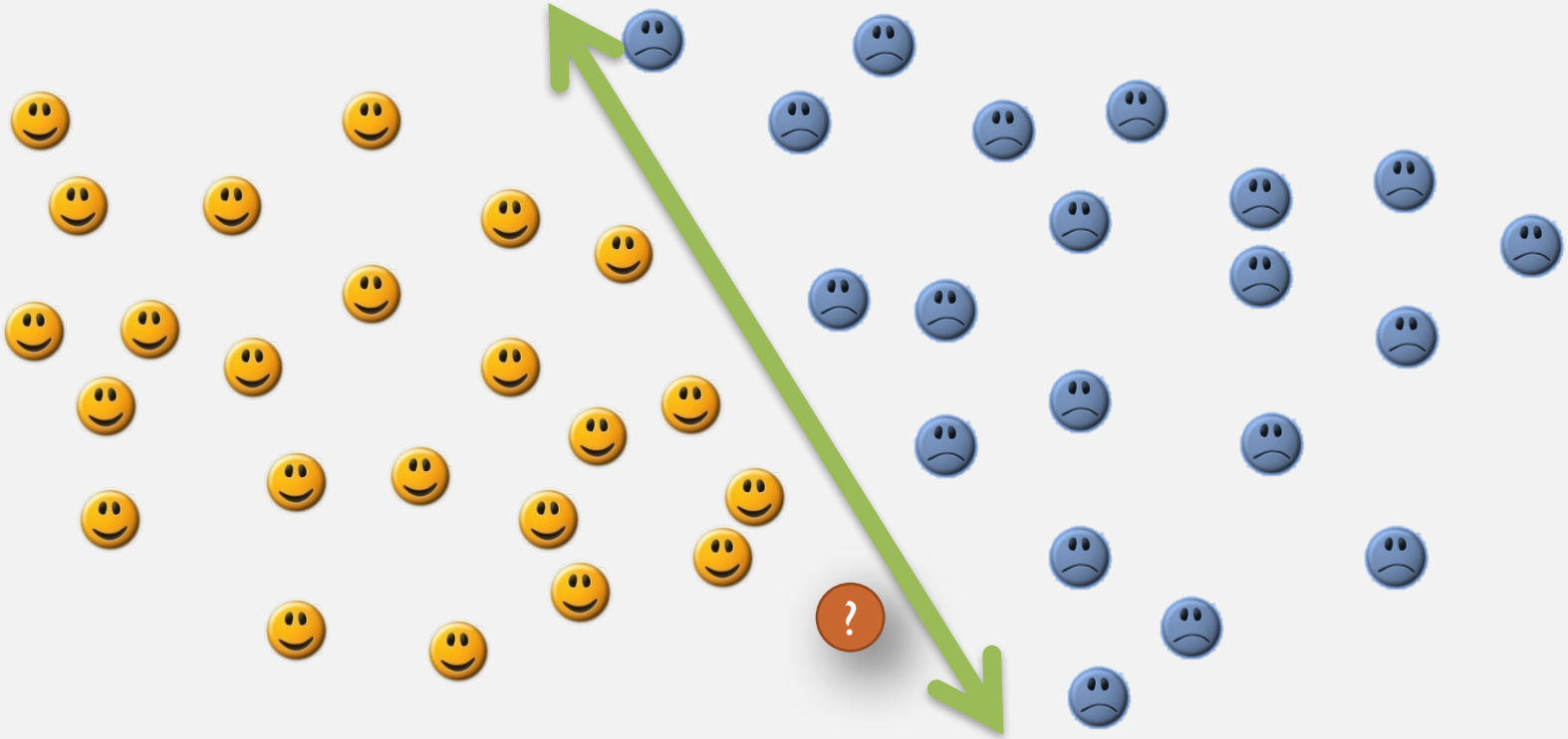


What if you're not a prolific, deceased, female, British author?

MACHINE LEARNING TO THE RESCUE

BINARY AND LINEARLY SEPARABLE

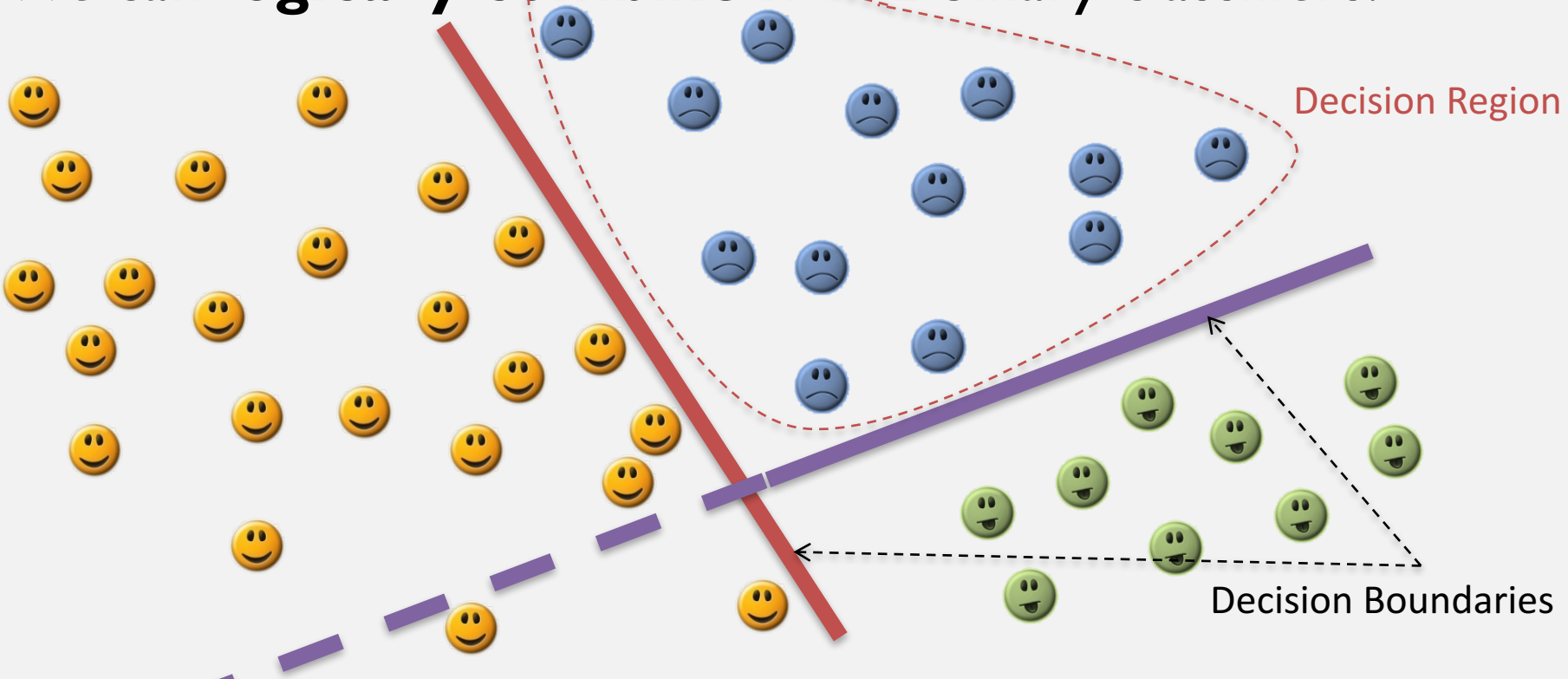
- Perhaps the easiest case.
  - Extends to dimensions $d \geq 3$, line becomes hyper-plane.



N-ARY AND LINEARLY SEPARABLE

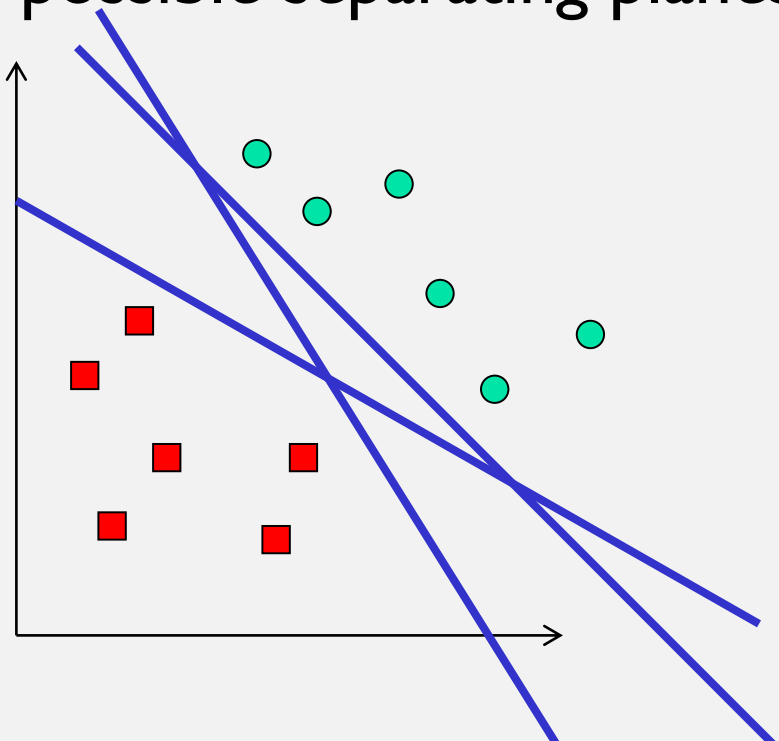
- A bit harder – random guessing might give $\frac{1}{N}$ accuracy

  We can **logically combine** $N - 1$ binary classifiers.



SUPPORT VECTOR MACHINES (SVMS) I

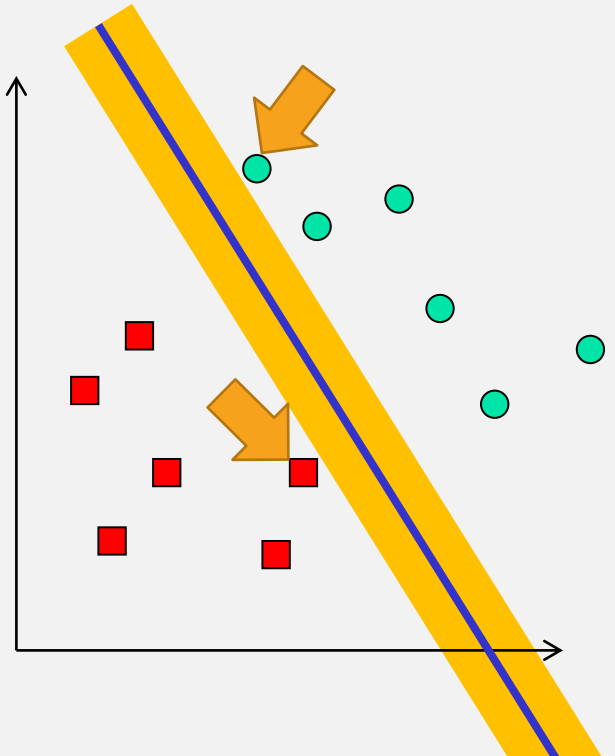
- In binary linear classification, two classes are assumed to be separable by a line (or plane). However, many possible separating planes might exist.



- Each of these blue lines separates the training data.
  - *Which line is the best?*

SUPPORT VECTOR MACHINES (SVMS) 2

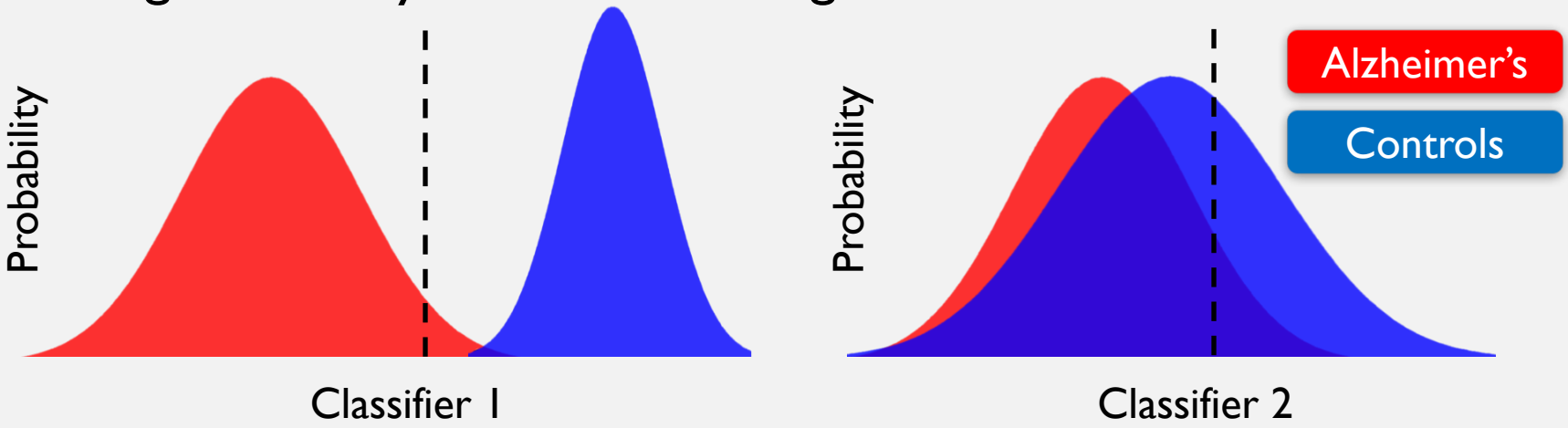
- The **margin** is the width by which the boundary could be **increased** before it hits a training datum.



- The **maximum margin linear classifier** is ∴ the linear classifier with the maximum margin.
- The **support vectors** (indicated) are those data points against which the margin is pressed.
- The bigger the margin – the less sensitive the boundary is to error.

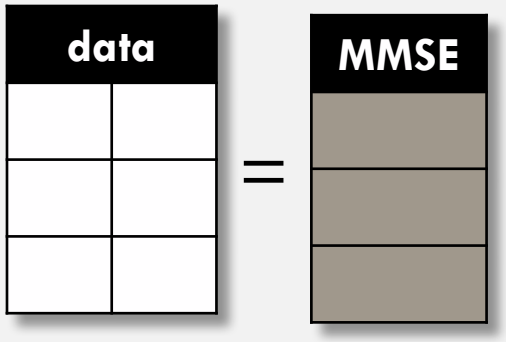
MANY DIFFERENT TYPES OF MACHINE LEARNING I

- SVMs are but one of a family of '**discriminative**' classifiers whose aim is to minimize error.
  - E.g., decision trees, many types of neural network, …
- Other '**generative**' classifiers learn *representations* of the phenomenon itself, in order to make a decision.
  - E.g., naïve Bayes, multilinear regression



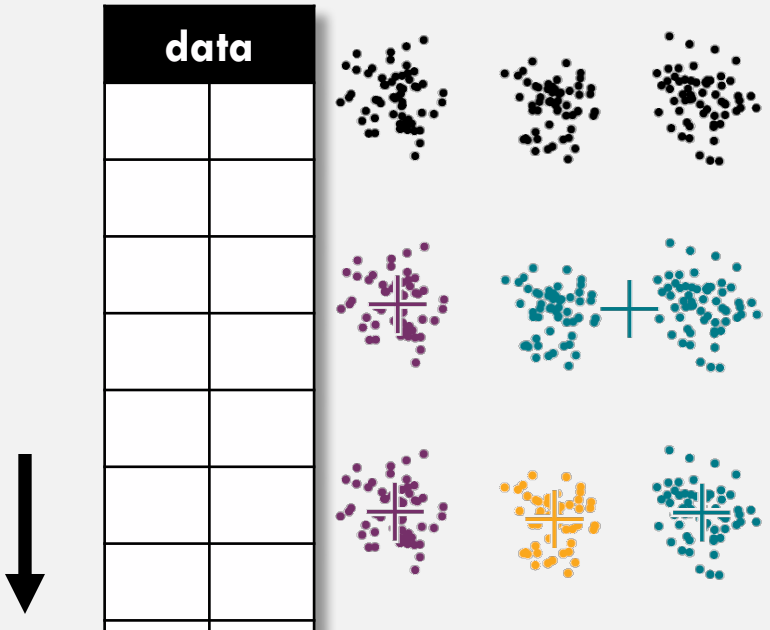
MANY DIFFERENT TYPES OF MACHINE LEARNING 2

- '**Supervised**' machine learning needs labeled data.



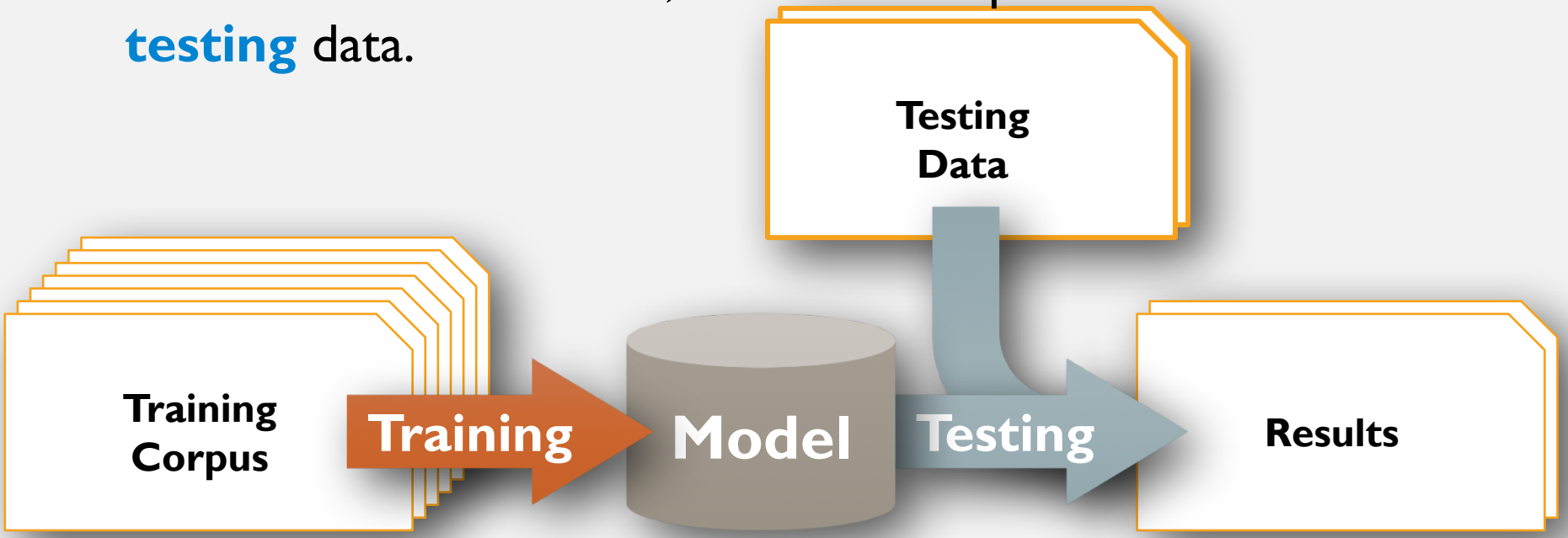
VS

- '**Unsupervised**' machine learning identifies patterns without getting explicit labels.



GENERAL PROCESS I

1. We gather a big and relevant **training** corpus.
2. We learn our **parameters** (e.g., probabilities) from that corpus to build our **model**.
3. Once that model is fixed, we use those probabilities to evaluate **testing** data.



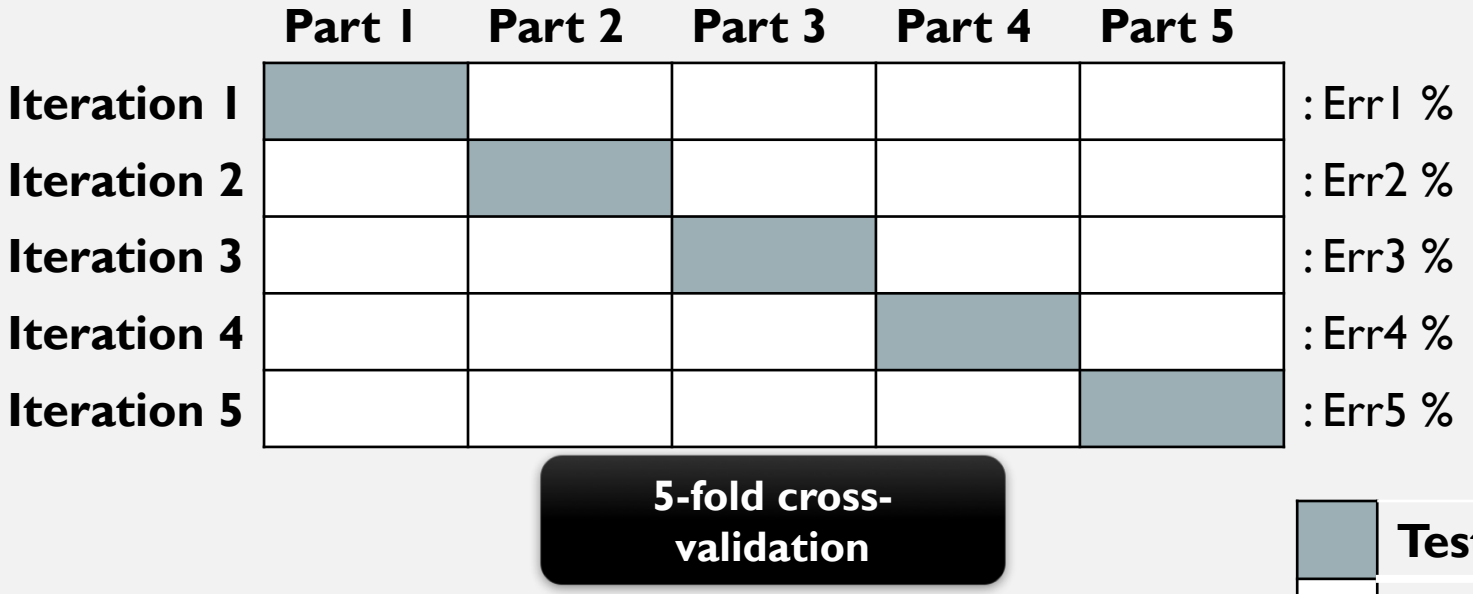
GENERAL PROCESS 2

- Often, **training data** consists of 80% to 90% of the available data.
  - Often, some subset of *this* is used as a **validation/ development set**.

- **Testing data** is ***not*** used for training but comes from the same *corpus*.
  - It often consists of the remaining 10% to 20% of the data.
  - Sometimes, it's important to **partition** speakers/writers so they **don't** appear in both training and testing.

BETTER PROCESS: *K*-FOLD CROSS-VALIDATION

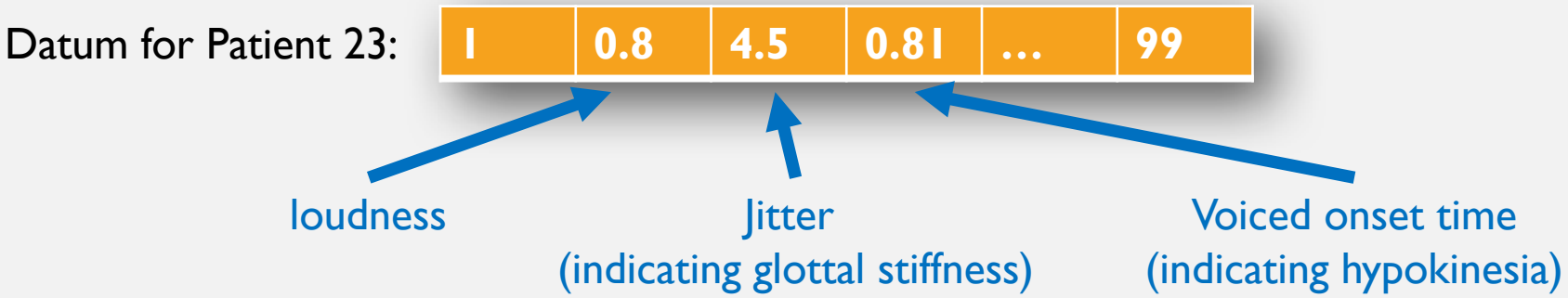
- ***K*-fold cross validation**: *n.* splitting all data into *K* **partitions** and iteratively testing on each after training on the rest (report means and variances).



WHAT IS A DATUM?

FEATURE VECTORS

- In general, each observation becomes a **vector** of numbers, each of which is a value of a particular descriptive **feature**.
- E.g., if you were to analyze someone's **voice** for signs of **Parkinson's disease**, a datum might represent a single utterance of them saying /ah/ for 5 seconds, from which you'd derive measures like:



CATEGORIES OF LINGUISTIC KNOWLEDGE

- **Phonology**: the study of patterns of speech sounds.
  e.g., "read" → /r iy d/
- **Morphology**: how words can be changed by inflection or derivation.
  e.g., "read", "reads", "reader", "reading", …
- **Syntax**: the ordering and structure between words and phrases (i.e., grammar).
  e.g., *NounPhrase → article adjective noun*
- **Semantics**: how meaning is created by words and phrases.
  e.g., "book" → 
- **Pragmatics**: the study of meaning in contexts.

WORDS (LEXICAL LEVEL) I

- How should we deal with these words?
  - *run* vs *runs* (verb conjugation)
  - *happy* vs *happily* (adjective vs. adverb)
  - *fra*$^{(l)}$*gment* vs *fragme*$^{(l)}$*nt* (spoken stress)
  - *realize* vs *realise* (spelling)
  - *We* vs *we* (capitalization)

- How do we count speech disfluencies?
    - e.g., *I* ***uh*** ***main-****mainly do data processing*
  - Answer: It depends on our task.

WORDS (LEXICAL LEVEL) 2

- Usually, we **preprocess** the data. Typically:
  - Convert all data to **lowercase.**
  - Remove "**punctuation**"!?
  - **Lemmatize** or **stem** each word
    - I.e., conflate inflected/derived words to a *stem* (root)
    - **Porter stemmer** is often the default. It applies about 60 rules to words, including:
      - Gets rid of plurals and -ed or -ing suffixes
      - Deals with suffixes, -full, -ness, etc.

WORDS (LEXICAL LEVEL) 4

- Next, we **extract** some features.
  E.g., for each utterance:
  - Count the number of **disfluencies**,
  - Count the number of **tokens**,
  - Measure vocabulary **richness**,
    - e.g., **Honoré** statistic:

$$\frac{100 \log N}{1 - \frac{V_1}{V}}$$

  - …

N	is the number of tokens
V	is the number of types
V_1	is the number of types occurring once.

WORDS (LEXICAL LEVEL) 5

- Some 'parts-of-speech':

Part of Speech	Description	Examples
Noun (NN)	is usually a **person**, **place**, **event**, or **entity**.	*chair, prescriptions, kidney, patient.*
Verb (VB)	is usually an **action** or **predicate**.	*do, prescribe, form-1.*
Adjective (JJ)	modifies a **noun** to further describe it.	*orange, rambling, disgusting.*
Adverb (RB)	modifies a **verb** to further describe it.	*tenderly, often*

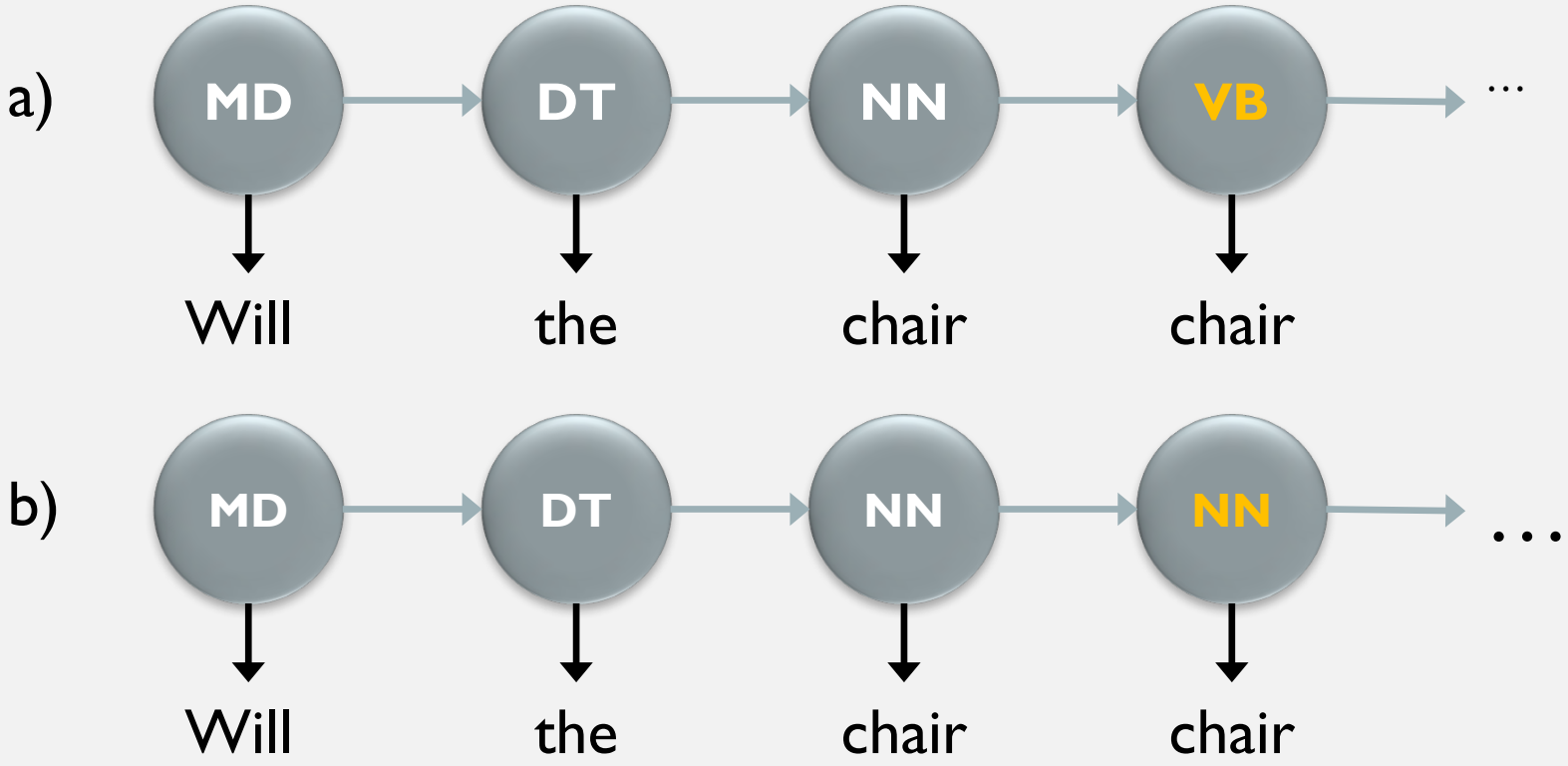
WORDS (LEXICAL LEVEL) 6

- Words can belong to many parts-of-speech.
  - E.g., *back:*
    - *The back/JJ door* (adjective)
    - *On its back/NN* (noun)
    - *Win the voters back/RB* (adverb)
    - *Promise to back/VB you in a fight* (verb)

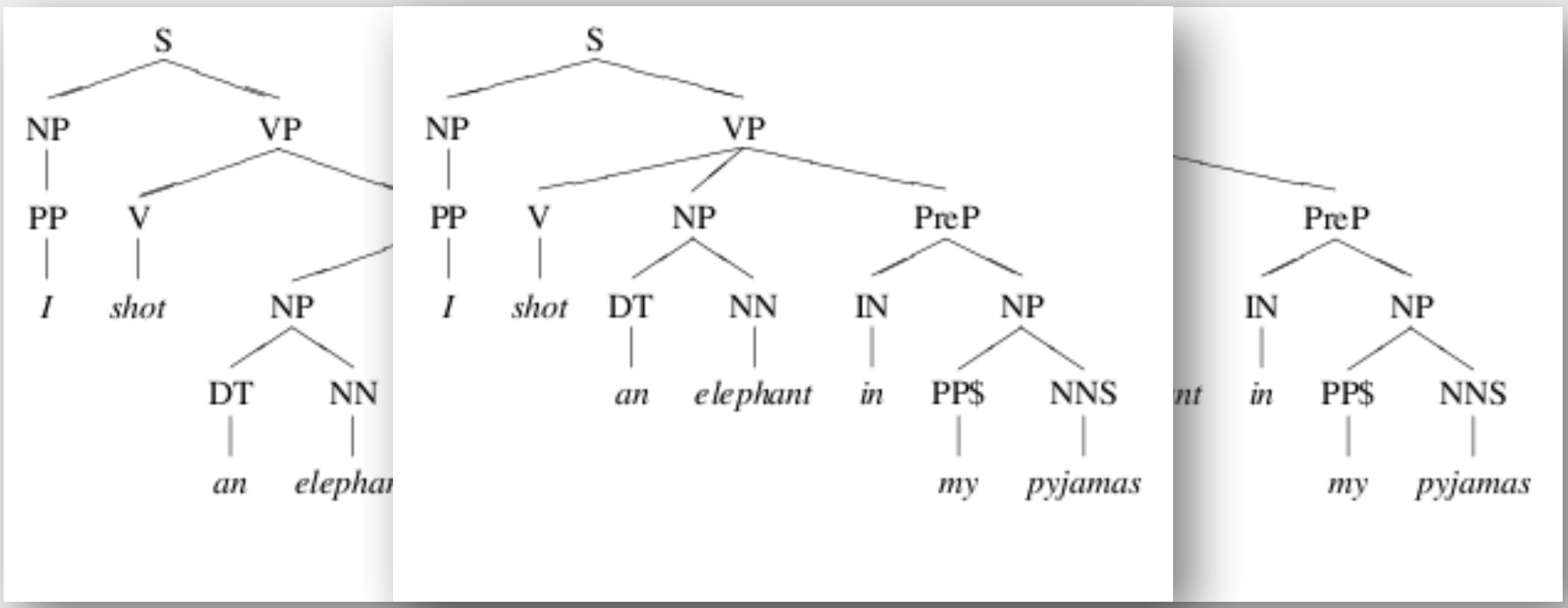
- We want to decide the **appropriate** tag given a particular sequence of tokens.

(ASIDE)
HIDDEN MARKOV MODELS

- *Will/MD the/DT* ***chair/NN*** ***chair/??*** *the/DT meeting/NN from/IN that/DT* ***chair/NN****?*



GRAMMAR I


S
NP
VP
PP
V
I
shot
NP
DT
NN
an
S
NP
VP
PP
V
NP
PrepP
I
shot
DT
NN
IN
NP
an
elephant
in
PP$
NNS
my
pyjamas
PrepP
IN
NP
in
PP$
NNS
my
pyjamas

GRAMMAR 2

- *Assuming* a parse is correct, we can
  - Count prepositional phrases (**PPs**)
  - Compute the **depth** of the tree (count the maximum length of ')' sequences),
  - …

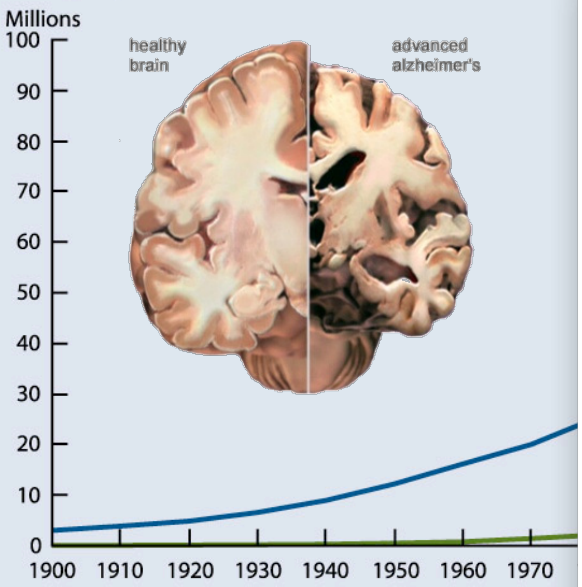
```
(ADJP
          (JJ kid)
          (JJ tryin)
          (S+VP
            (TO to)
            (VP  (VB get)
                 (PP (IN into)
                     (NP (DT the) (JJ cookie) (NN jar))))))
```

6

YOUR PROJECT

THE RISING TIDE OF DEMENTIA

Number of people age 65 and over, by age group, selected years 1900–2006 and projected 2010–2050



Note: Data for 2010–2050 are projections of the population.
Reference population: These data refer to the resident population.
Source: U.S. Census Bureau, Decennial Census, Population Estimates and Projections.

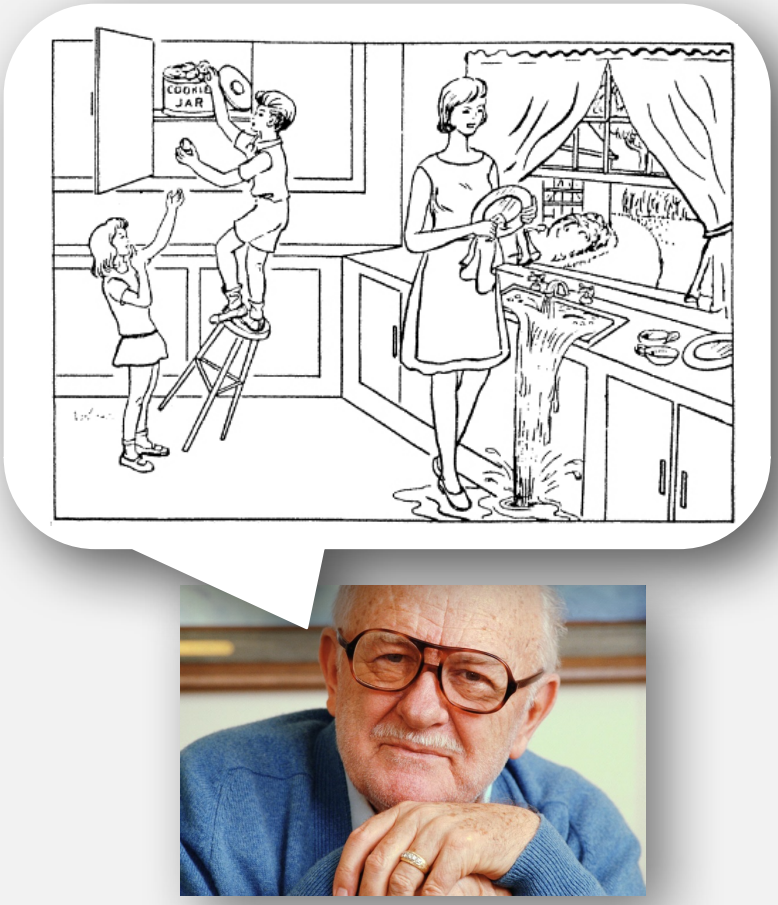
Mini-Mental State Examination (MMSE)

Patient's Name: _____ Date: _____

Instructions: Score one point for each correct response within each question or activity.

Maximum Score	Patient's Score	Questions
5		"What is the year? Season? Date? Day? Month?"
5		"Where are we now? State? County? Town/city? Hospital? Floor?"
3		The examiner names three unrelated objects clearly and slowly, then the instructor asks the patient to name all three of them. The patient's response is used for scoring. The examiner repeats them until patient learns all of them, if possible.
5		"I would like you to count backward from 100 by sevens." (93, 86, 79, 72, 65, …) Alternative: "Spell WORLD backwards." (D-L-R-O-W)

ASSESSING ALZHEIMER'S AUTOMATICALLY

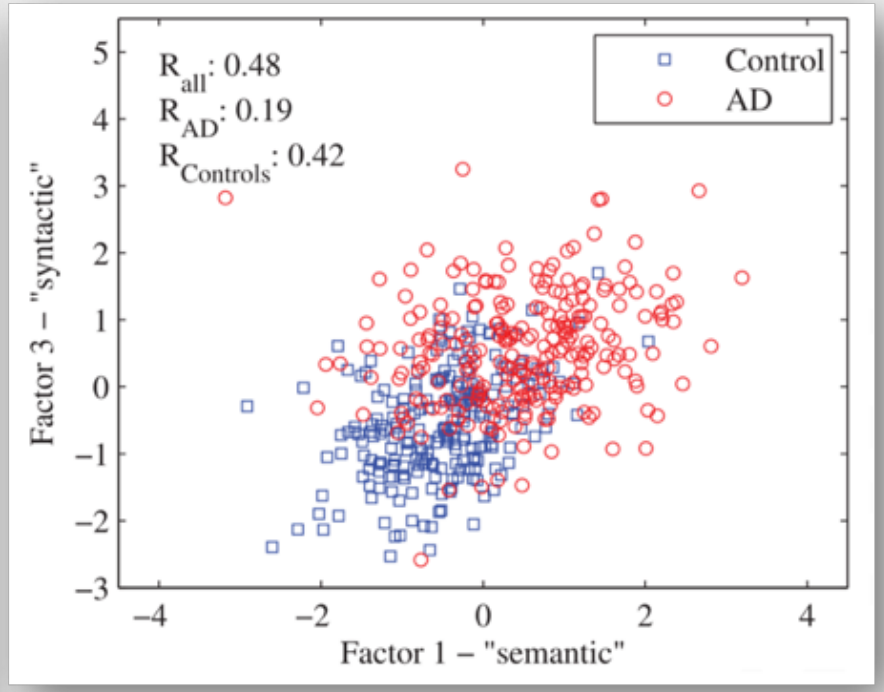


- A task that can be done in less than a minute, on the couch.

- **DementiaBank:**
  240 samples from167 people with AD,
  233 samples from 97 controls.
  - Free-form descriptions of "Cookie Theft" (incl. audio)
  - Transcribed and annotated, e.g., with filled pauses, paraphasias, and unintelligible words.
  - **Mini-mental state exam (MMSE)**

ASSESSING ALZHEIMER'S AUTOMATICALLY

Lexical	Noun-to-pronoun ratios; Avg. word length; # demonstratives; Familiarity; Honoré statistic
Syntactic	Parse tree depth; VP → VPG; VP → AUX VP; Coordinate conjunctions; Mean clause length
Acoustic	Phonation rate; Mean F2; Mean RPDE; Mean power; Pause::word ratio



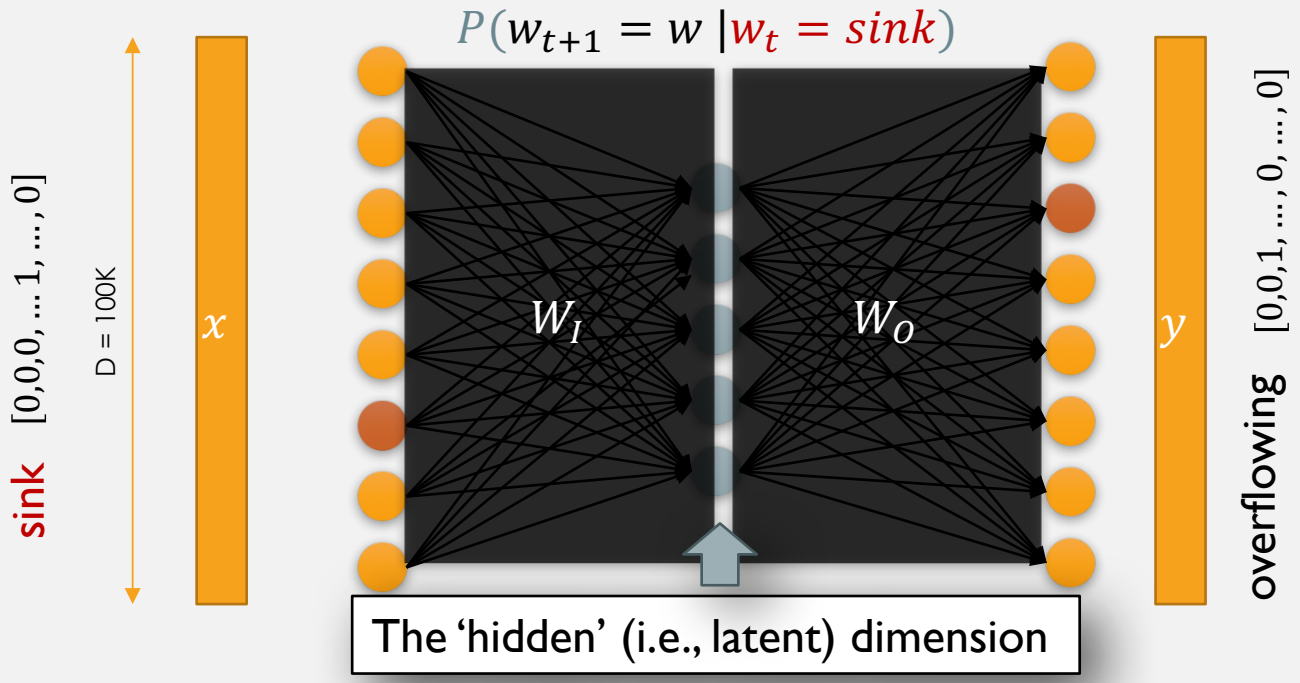
State-of-the-art accuracy: 85% - 92%

How do you measure semantics?

QUICK COMMENT ON NEURAL NETWORKS

(this is just for your general interest)

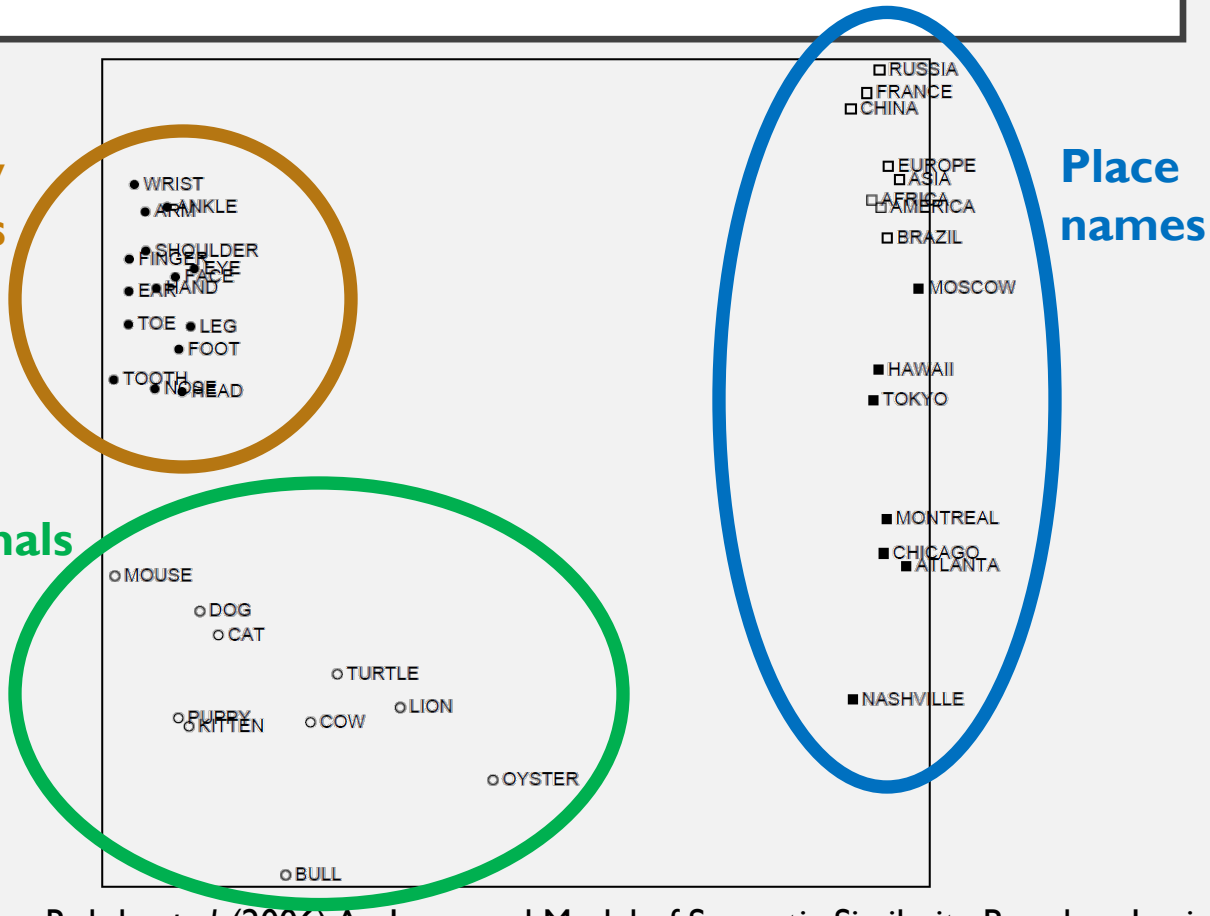
NEURAL NETWORKS



We i) '**plug in**' *each* word *in sequence*, ii) **perform** matrix multiplication, iii) **compare** the result to the next word, and iv) **propagate** the error back through the weights.

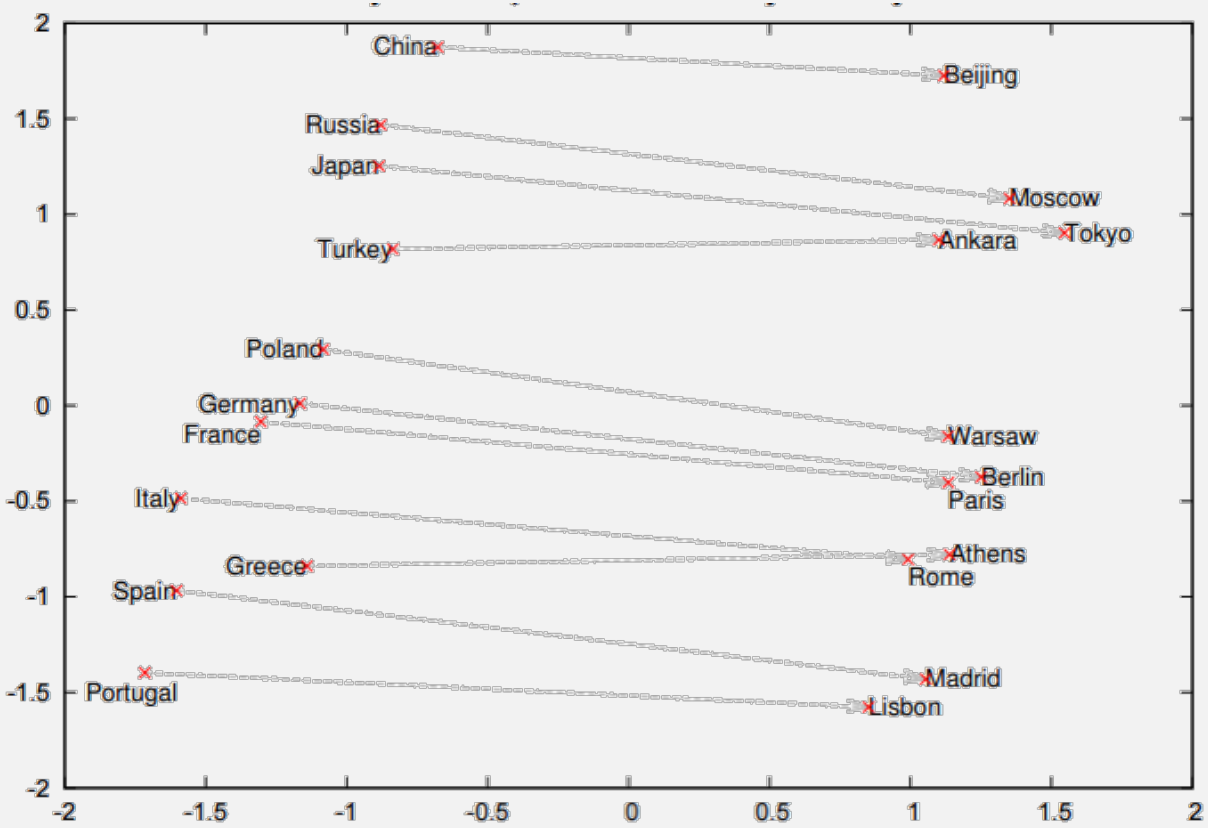
PROJECTING WORDS TO HIDDEN DIMENSIONS

for a given word x, xW_I gives a 'latent' vector



Rohde *et al.* (2006) An Improved Model of Semantic Similarity Based on Lexical Co-Occurrence. *Communications of the ACM* **8**:627-633.

REGULARITIES IN WORD-VECTOR SPACE



Trained on the Google news corpus with over 300 billion words.

REGULARITIES IN WORD-VECTOR SPACE

Expression	Nearest token
Paris – France + Italy	Rome
Bigger – big + cold	Colder
Sushi – Japan + Germany	bratwurst
Cu – copper + gold	Au
Windows – Microsoft + Google	Android

Analogies: apple:apples :: octopus:octopodes
Hypernymy: shirt:clothing :: chair:furniture

QUICK COMMENT ON TEXT INFORMATICS GENERALLY

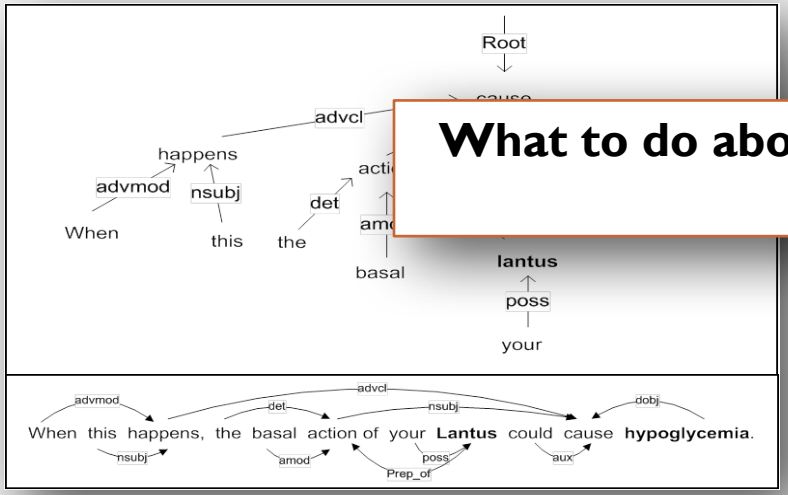
(this is just for your general interest)

TEXT INFORMATICS IN MEDICINE

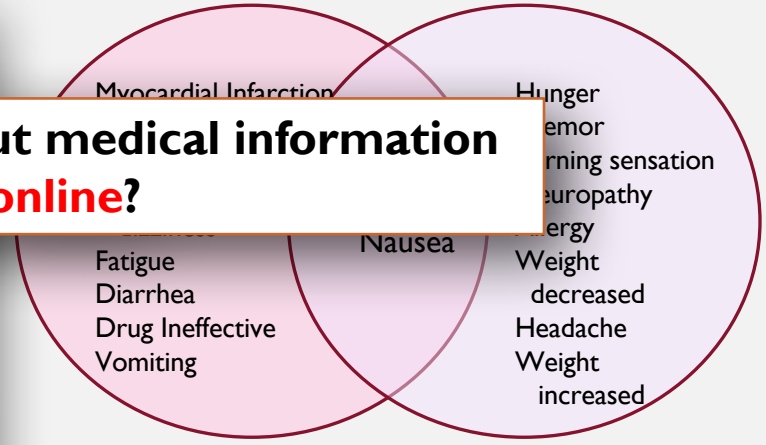
- Analyzing things doctors say.
  - E.g., differential diagnosis from the medical record.
- Analyzing things doctors are told.
  - E.g., knowledge discovery from medical texts.
- Analyzing things the public say.

FORMAL AND INFORMAL LANGUAGE IN HEALTHCARE

- People generally *describe things differently* than doctors.
  - They also *describe different things*.



What to do about medical information online?



FAERS

Forums

Liu, X., & Chen, H. (2015). Identifying adverse drug events from patient social media: A case study for diabetes. IEEE Intelligent Systems, 30(3):44–51. *(i.e., not us)*

POST-TRUTH, ONLINE

Vaccines DO Cause Autism-Undeniable Scientific Proof – Anti ... ✗
https://avscientificsupportarsenal.wordpress.com/.../**vaccines**-do-cause-**autism**-undenia...
Apr 29, 2015 - There is absolutely undeniable scientific proof that **vaccines** cause **autism**. There is no question! Case closed! Game over! The people and the ...
You visited this page.

Autism-Vaccine Link: Evidence Doesn't Dispel Doubts - WebMD ✓
www.webmd.com/brain/**autism**/searching-for-answers/**vaccines-autism**
Many major medical groups say **vaccines** don't cause **autism**. Many parents say they do. So who's right?

Vaccines and Autism: A Tale of Shifting Hypotheses ✓
cid.oxfordjournals.org/content/48/4/456.full
by S Plotkin - 2009 - Cited by 189 - Related articles
Three specific hypotheses have been proposed: (1) the combination measles-mumps-rubella **vaccine** causes **autism** by damaging the intestinal lining, which allows the entrance of encephalopathic proteins; (2) thimerosal, an ethylmercury-containing preservative in some **vaccines**, is toxic to the central nervous system; and (3 ...

Vaccines and autism: Separating fact from fiction | BabyCenter ✓
www.babycenter.com › Baby › Baby Development › Autism
Did a preservative in children's **vaccines** cause a rise in **autism** rates? We examine the evidence.

MMR Vaccine and Autism - Immunize Canada ✓
www.immunize.ca › Home › Publications and Resources › Questions & misconceptions
Some speculation has tried to link thimerosal in the MMR **vaccine** to **autism**, but the MMR **vaccine** routinely used in Canada has never contained thimerosal.

SUMMARY

- To do machine learning,
  - You separate data so you can train models and then test them.
  - You describe each datum by a vector of features, e.g., the richness of the vocabulary, the ratio of nouns to pronouns, etc.
  - You choose a model type, e.g., support vector machines.
- We can identify Alzheimer's disease by listening to short snippets of picture descriptions.

Thank you!