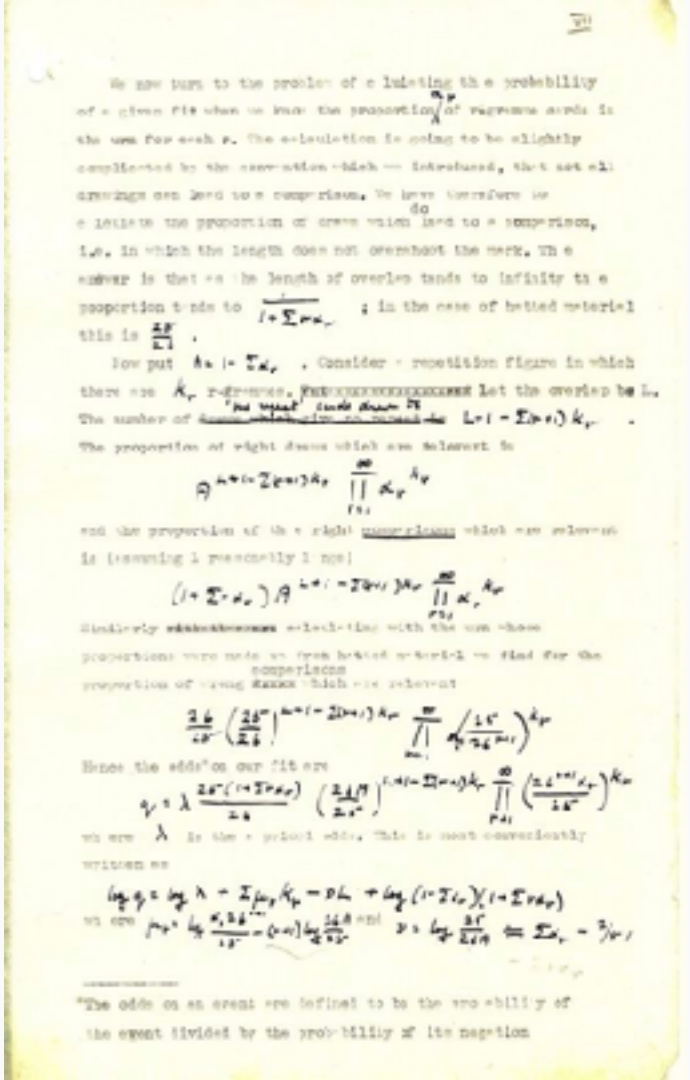
Practical Tips for Research Writing

Kris Sankaran
(Mila -> UW Madison)

July 3, 2020 — AMMI Tea Talks

Research (writing) is hard

- Writing is too hard if you think of it as one big task
  - How can we break it into more manageable pieces?
- Each piece involves skills that you can improve over time
  - Certain reading and brainstorming habits help
- **Goal: Nothing fancy, just lots of practical tricks**



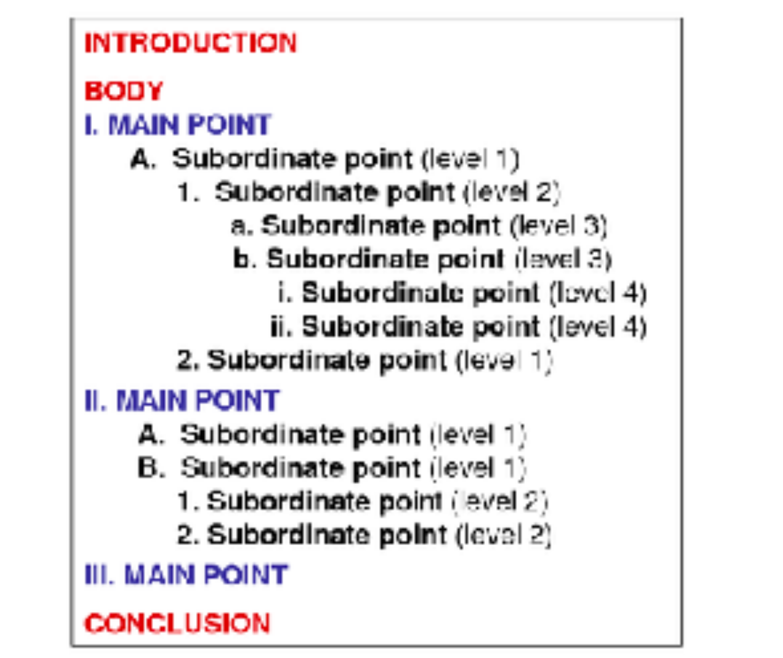
Part 1: The writing process

Stages of writing: Outline

- Overall scaffolding for your piece

- Approaches,
  - **Indented list, with topics / subtopics**
  - Mind-Map

- Start this early: a good outline can tell you want experiments you'd need to run



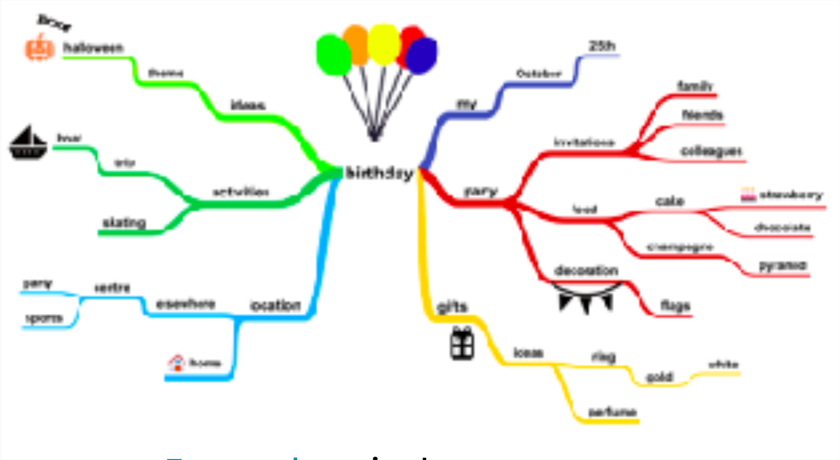
Generic [outline structure].

Stages of writing: Outline

- Overall scaffolding for your piece

- Approaches,
    - Indented list, with topics / subtopics
    - **Mind-Map**

- Start this early: a good outline can tell you want experiments you'd need to run

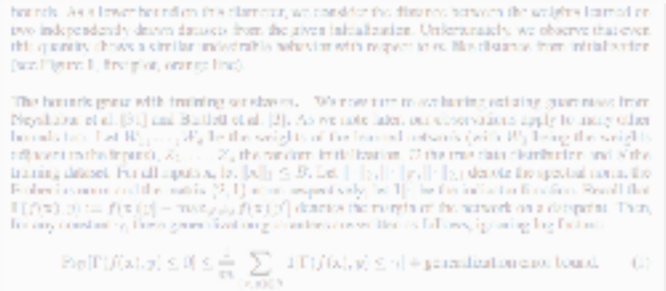


Example mind-map.

Stages of writing: Outline

- **Gather all your primary sources**
  - Relevant citations
  - Algorithmic details
  - Experimental results to share
- Reverse outlining target conference
  - Take existing papers, and imagine the outline that they used

Purpose: Decouple analytical from imaginative thinking.

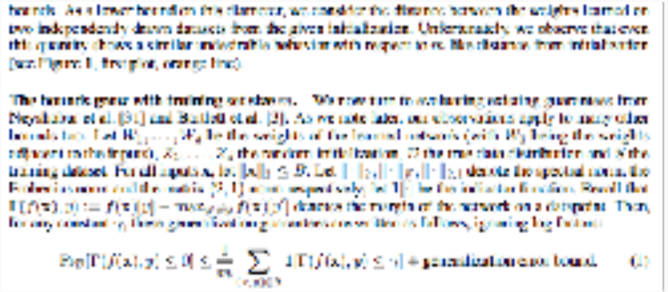


```
..
III. Existing bounds vs. training set size
  A. Overview of section
  B. Setup notation
  C. Fact: Norm grows with training set size
    1. Figure giving evidence
    2. Interpretation, why this is contrary to usual thinking
  D. Fact: Bounds grow with training set size
```

Stages of writing: Outline

Purpose: Decouple analytical from imaginative thinking.

- Gather all your primary sources
  - Relevant citations
  - Algorithmic details
  - Experimental results to share
- **Reverse outlining target conference**
  - Take existing papers, and imagine the outline that they used



```
..
III. Existing bounds vs. training set size
  A. Overview of section
  B. Setup notation
  C. Fact: Norm grows with training set size
    1. Figure giving evidence
    2. Interpretation, why this is contrary to usual thinking
  D. Fact: Bounds grow with training set size
```

Stages of writing: Draft

- Deliberately write quickly and roughly
  - Avoid self-editing!

- Time yourself, make it feel like a high school exam. 2 hours is a good limit.

- ‘tk’ trick: If you don’t know what to put somewhere, just put ‘tk’

I use the outline vs. draft breakdown in almost everything I write, from emails to lectures to reviews to papers.

Stages of writing: Revise

- Check for the curse of knowledge and give examples
  - Where might the reader get tripped up?
- Make sure your contributions are obvious
  - You really have to hit them over the head
- Your goal isn't to sound smart, it's to make someone understand. Use words that I know.

A good conference reviewer will (1) summarize and (2) identify contributions of your paper. Pretend you are a reviewer.

Preparing Figures

Appendance

- **Aim for a high data-to-ink ratio**
- Make sure labels are legible
- Try to make self-explanatory

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$$

$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$

$= 1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$

Captions

- First describe: explain objective facts
- Then provide your interpretation

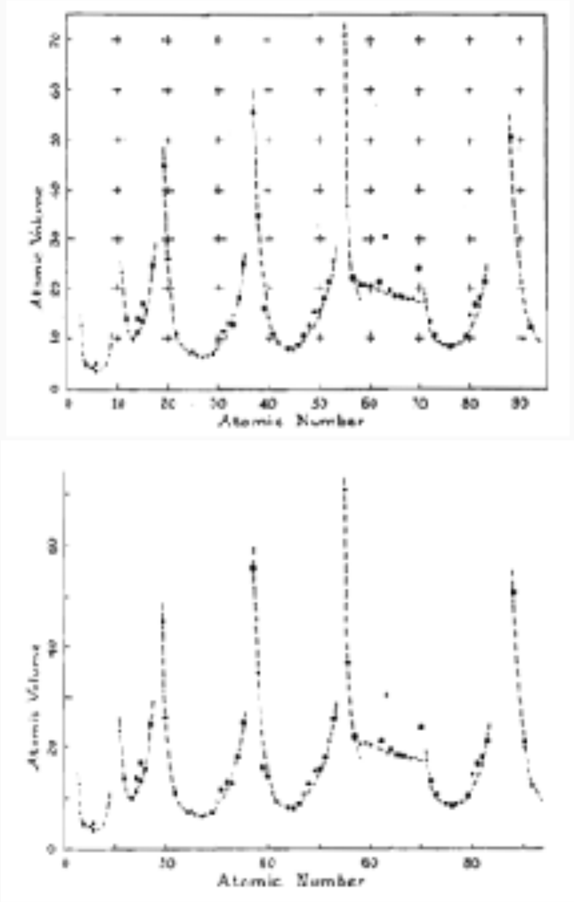
Preparing Figures

Appearance

- **Aim for a high data-to-ink ratio**
- Make sure labels are legible
- Try to make self-explanatory

Captions

- First describe: explain objective facts
- Then provide your interpretation



[Example](#) from Tufte's book.

Preparing Figures

Appearance

- Aim for a high data-to-ink ratio
- Make sure labels are legible
- Try to make self-explanatory

Captions

- **First *describe*: explain objective facts.**
- Then *interpret*: What have we learned?

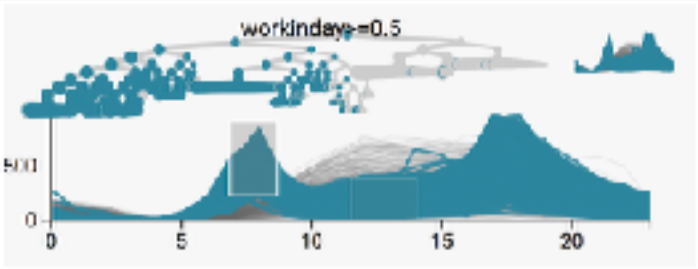


Figure 13: The two peaks at rush hour distinguish weekday series from the rest, through the timebox tree view. The display is the same type of timebox tree view introduced in Figure 1, but applied to the bikesharing data, where the time axis represents the time of day and the y-axis gives bikesharing demand. Each series is the bikesharing demand for a single day, over the course of two years. The tree now corresponds to the regression tree generated by predicting demand at 8am using supplementary data. Two brushes are introduced to highlight the double peaks corresponding to rush hours on weekdays. We see that although hierarchical structure was not present immediately in the bikesharing data, it is useful to introduce and interpret such structure by combining regression and visualization methodology.

A very long caption from my tree visualization paper.

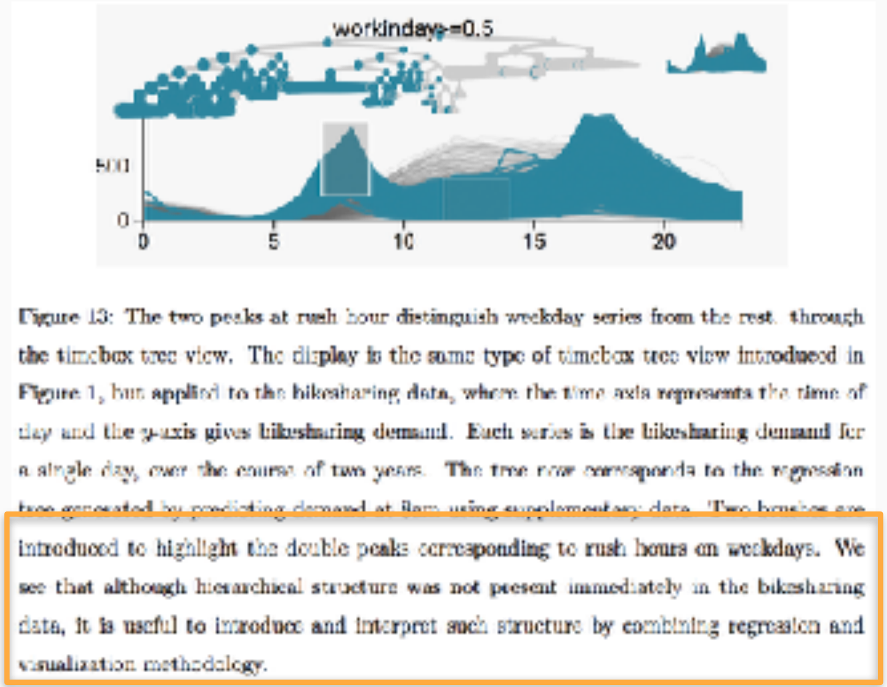
Preparing Figures

Appearance

- Aim for a high data-to-ink ratio
- Make sure labels are legible
- Try to make self-explanatory

Captions

- First *describe*: explain objective facts.
- **Then *interpret*: What have we learned?**



A very long caption from my tree visualization paper.

Part 2: The reading process

Reading

Understanding

- Find concrete examples
- Draw pictures
- Articulate what you don’t understand (a definition, an argument, …)

Engaging

- Summarize contributions. Does it live up to claims?
- Reconstruct authors’ thought process.
- Can you apply it to your problems?

Reading

Understanding

- **Find concrete examples**
- Draw pictures
- Articulate what you don't understand (a definition, an argument, ...)

Engaging

- Summarize contributions. Does it live up to claims?
- Reconstruct authors' thought process.
- Can you apply it to your problems?

and θ_j^M). To rate an item, a user first draws a topic z_{uj}^U from his distribution, representing, for example, his mood at the time of rating (in the mood for romance vs. comedy), and the item draws a topic z_{uj}^M from its distribution, representing, for example, the context under which it is being rated (in a theater on opening night vs. in a high-school classroom). The

Sometimes, the authors give us examples, (like this one from *Mixed-Membership Matrix Factorization*), other times we have to make one up ourselves.

Reading

Understanding

- **Find concrete examples**
- Draw pictures
- Articulate what you don't understand (a definition, an argument, ...)

Engaging

- Summarize contributions. Does it live up to claims?
- Reconstruct authors' thought process.
- Can you apply it to your problems?

bounds. As a lower bound on this diameter, we consider the distance between the weights learned on two independently drawn datasets from the given initialization. Unfortunately, we observe that even this quantity shows a similar undesirable behavior with respect to m like distance from initialization (see Figure 1, first plot, orange line).

The bounds grow with training set size m. We now turn to evaluating existing guarantees from Neyshabur et al. [31] and Bartlett et al. [3]. As we note later, our observations apply to many other bounds too. Let $W_1, \ldots, W_d$ be the weights of the learned network (with W_1 being the weights adjacent to the inputs), $Z_1, \ldots, Z_d$ the random initialization, $\mathcal{D}$ the true data distribution and S the training dataset. For all inputs $\mathbf{x}$, let $\|\mathbf{x}\|_2 \leq B$. Let $\|\cdot\|_2, \|\cdot\|_F, \|\cdot\|_{2,1}$ denote the spectral norm, the Frobenius norm and the matrix $(2,1)$-norm respectively; let $\mathbf{1}[\cdot]$ be the indicator function. Recall that $\Gamma(f(\mathbf{x}), y) := f(\mathbf{x})[y] - \max_{y' \neq y} f(\mathbf{x})[y']$ denotes the margin of the network on a datapoint. Then, for any constant γ, these generalization guarantees are written as follows, ignoring log factors:

$$\Pr_{\mathcal{D}}[\Gamma(f(\mathbf{x}), y) \leq 0] \leq \frac{1}{m} \sum_{(x,y) \in S} \mathbf{1}[\Gamma(f(\mathbf{x}), y) \leq \gamma] + \text{generalization error bound.} \qquad (1)$$

What are some special cases? What if x is one-dimensional? What happens when m = 0, 1, or infinity?

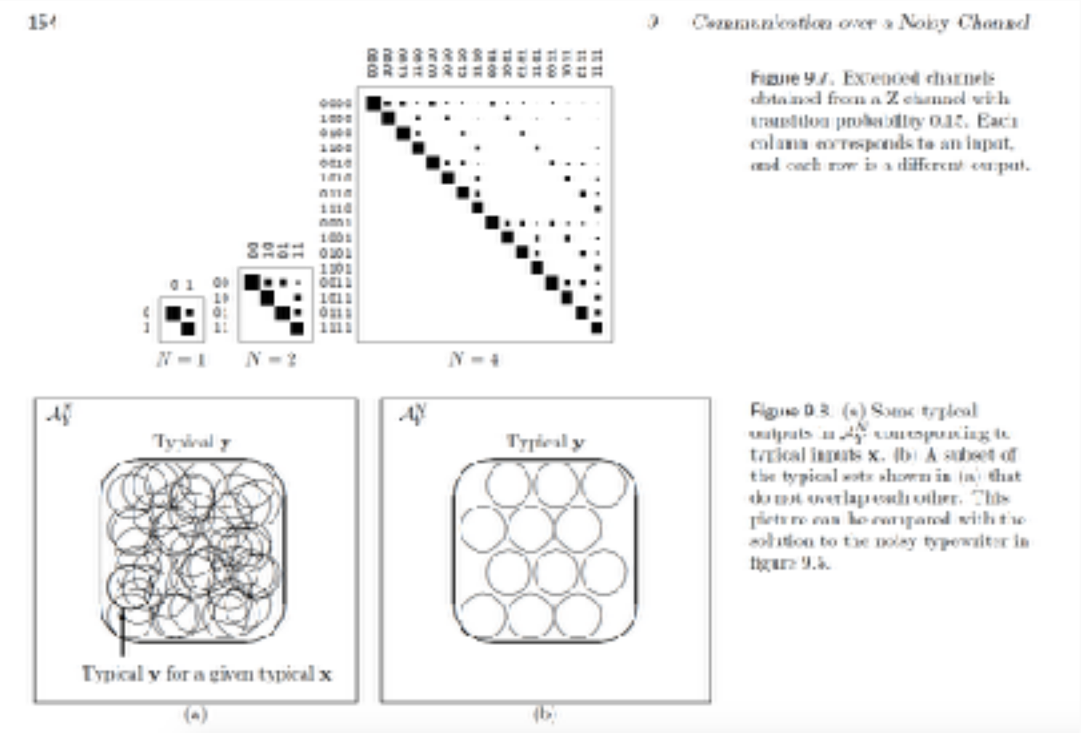
Reading

Understanding

- Find concrete examples
- **Draw pictures**
- Articulate what you don't understand (a definition, an argument, ...)

Engaging

- Summarize contributions. Does it live up to claims?
- Reconstruct authors' thought process.
- Can you apply it to your problems?



(example from David Mackay's Info Theory book, free online.)

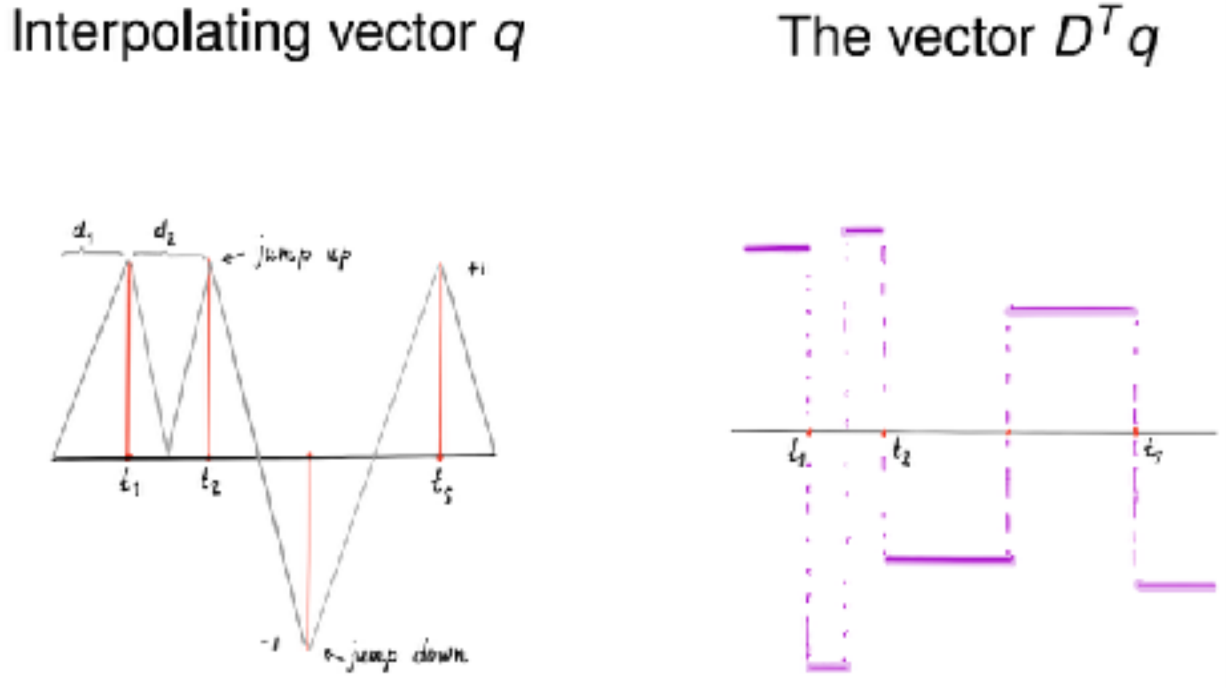
Reading

Understanding

- Find concrete examples
- **Draw pictures**
- Articulate what you don't understand (a definition, an argument, ...)

Engaging

- Summarize contributions. Does it live up to claims?
- Reconstruct authors' thought process.
- Can you apply it to your problems?



Sara van der Geer gave amazing illustrations in her talk... not all speakers will do this for you!

Exercise: What papers do you like / hate?

- Read 6 papers any conference proceedings in a topic you know
- Grade each paper from A+ to F
- *« grades should be childishly selfish and impudent measures of your own joy or lack of it »*
- **Only rule: Cannot give all papers the same grade.**

(inspired by Kurt Vonnegut's Final Term assignment)

Exercise: What papers do you like / hate?

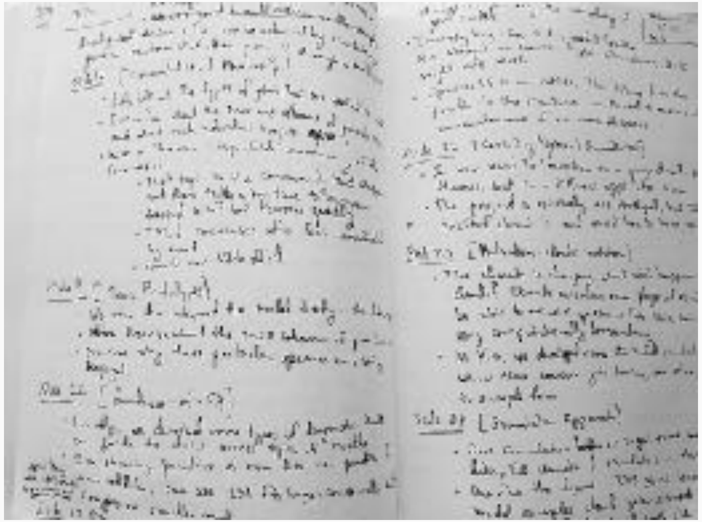
+	-
Clear motivation	Unclear what the goal is
Easy to understand	Unsure what was actually done
Prior work vs. contribution clear	Unclear what’s different from before
Interesting dataset / application	Overdone problem area
Good notation and examples	Terrible notation, no analogies
Sounds like a friendly person	Really obnoxious

Some ideas to get you started with the paper grading exercise.

Part 3: Research community

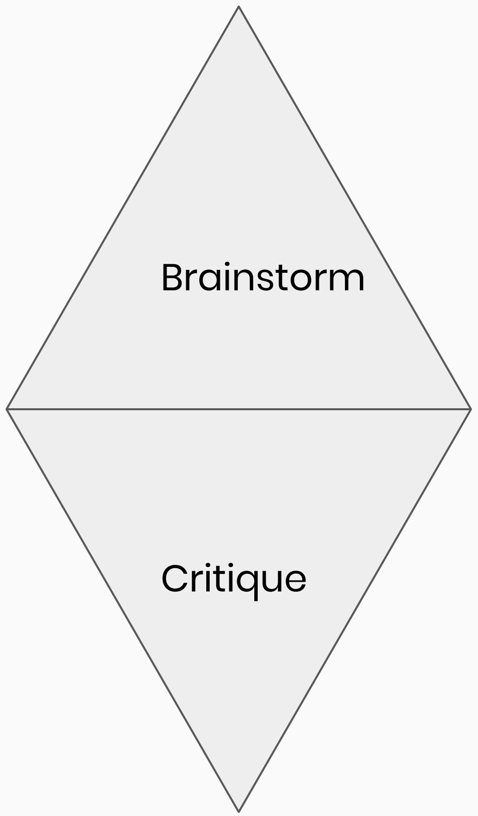
Lab Notebook

- Fragments of
  - Summarized readings
  - Examples you like
  - Problems you care about
  - Ideas you come up with
- Write informally (letters to self)
- Review periodically (informs paper writing)



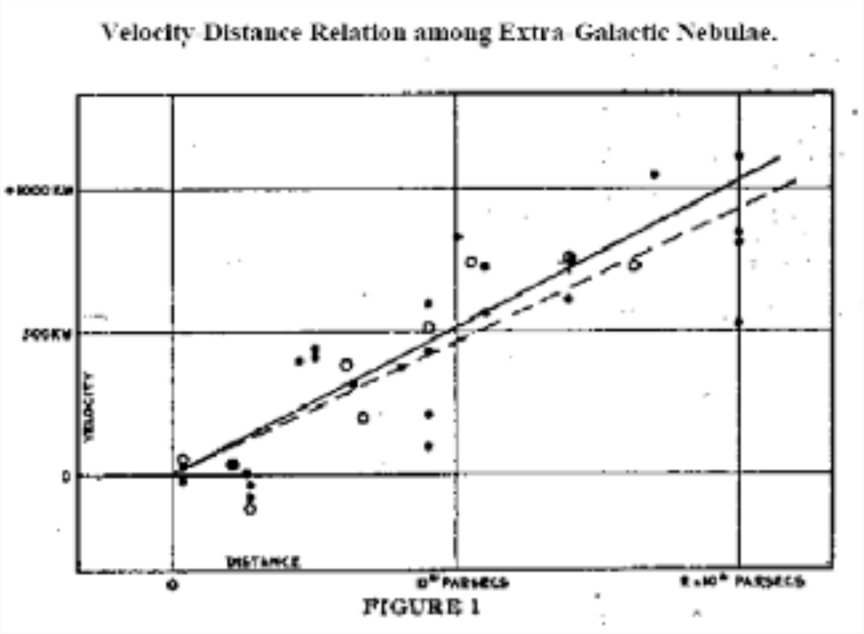
What to do at whiteboards

- Two modes: **Brainstorm and Critique**
    - Need to be able to tolerate ambiguity
- Imagine possible experiment outcomes
- Try articulating ideas as they come up



What to do at whiteboards

- Two modes: Brainstorm and Critique
    - Need to be able to tolerate ambiguity
- Imagine possible **experiment outcomes**
- Try articulating ideas as they come up



What figures will you draw?
What conclusions might they suggest?

What to do at whiteboards

- Two modes: Brainstorm and Critique
    - Need to be able to tolerate ambiguity
- Imagine possible experiment outcomes
- Try **articulating ideas** as they come up

Ideas should lead to falsifiable claims.

“New Model will learn from imperfect labels, while Old Model will not.” → Falsifiable
“New Model is really awesome” → Not falsifiable

Writing and Sharing

- **Sketch papers before ideas are fully-formed**
- Get feedback from others
- Research is a conversation, papers are the units of exchange

Gap: Thing we don’t know / don’t know how to do easily
Idea: What should we do about it?
Experimentation: How will you evaluate? What figures to make?
Interpretation: What can you conclude, and what’s next?

Writing and Sharing

- **Sketch papers before ideas are fully-formed**
- Get feedback from others
- Research is a conversation, papers are the units of exchange

1 Introduction

There are two basic ways to implement function application in a higher-order language, when the function is unknown: the *push/enter* model or the *eval/apply* model [11]. To illustrate the difference, consider the higher-order function **zipWith**, which zips together two lists, using a function **k** to combine corresponding list elements:

```
zipWith :: (a->b->c) -> [a] -> [b] -> [c]
zipWith k []     []     = []
zipWith k (x:xs) (y:ys) = k x y : zipWith xs ys
```

Here **k** is an *unknown function*, passed as an argument; global flow analysis aside, the compiler does not know what function **k** is bound to. How should the compiler deal with the call **k x y** in the body of **zipWith**? It can't blithely apply **k** to two arguments, because

Use an example to introduce the problem

Writing and Sharing

- Sketch papers before ideas are fully-formed
- **Get feedback from others**
- Research is a conversation, papers are the units of exchange

> to be working in. At this point—or even earlier—it's important to get plugged into the Secret Paper Passing Network. This informal organization is where all

From the "How to do research at the MIT AI Lab" Working Paper (1988)

Writing and Sharing

- Sketch papers before ideas are fully-formed
- Get feedback from others
- **Research is a conversation, papers are the units of exchange**

I think of science as a conversation that is carried out through paper-sized units. Any single paper can only do so much – it must have finite scope, so that the work behind it can be done in finite time and described in a finite number of pages. There is a limit on how much framing and explanation can fit into any paper. Supplemental materials can expand that scope somewhat, but even without explicit length limits for them there must still be a boundary.

Worthwhile problems

“No problem is too small or too trivial if we can really do something about it.”

From a wonderful letter by Richard Feynman to Koichi Mano.

Resources

- How to do research at the MIT AI Lab
- Graduate Study in the Computer and Mathematical Sciences: A Survival Guide
- How to have a bad career in research
- How to write a great research paper
- Stanford Reading Group Tips
- Ravi Vakil's advice
- J. Michael Steele's rants
- You and your research
- Worthwhile Problems