Methodology and Tools for Research:

Reading scientific material

Yannick Prié
Polytech Nantes, University of Nantes
Master DMKM, 2015-2016

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Objectives of this course

- How to read a paper?
- Various kinds of validations
- Kinds of writings related to reading
  - Active reading and annotations
  - Reading notes
  - Synthesis
  - Reviews

Outline

- Why would you read a paper?
- What are papers?
- How to read a paper?
- Results and evaluation
- Writing while reading

Outline

- **Why would you read a paper?**
- What are papers?
- How to read a paper?
- Results and evaluation
- Writing while reading

Why would you read a paper? (1)

- To know what it is about
  - superficial reading, skimming
- To fully assimilate its content
  - deep reading
- To find an information you know it contains
  - a sentence to quote, a figure, a reference

Why would you read a paper? (2)

- To write a reading note about it
  - private evaluation
- To review it
  - for a conference or journal: formal evaluation
  - for a colleague
- To include it in a synthesis note
  - survey
  - literature review
- To include it as a reference
  - in another paper, PhD thesis, grant proposal...
- To get inspiration

To write!

Outline

- Why would you read a paper?
- **What are papers?**
- How to read a paper?
- Results and evaluation
- Writing while reading

Different kinds of papers

- Theoretical papers
  - Theorems, proofs, concepts
  - e.g. problem solving, algorithm complexity, formal semantics of a language
- Empirical papers
  - Proposal and experimental evaluation
  - Description and analysis
  - e.g. of architecture, algorithm, model, software, usages
- Surveys
  - Exhaustive state of the art on a subject
  - e.g. sequence mining
  - (in other disciplines: meta-review)
- ...

What's in a paper?

Title

Author(s)

Abstract

Keywords

Introduction — Problem statement

Body — Claim; Elements of validation; Discussion; Comparison to the literature

Conclusion and future work

References

For a classical experimental paper

- Title
- Author(s)
- Abstract
- Keywords
- Introduction
- **Related works**
- **System description**
- **Evaluation**
- **Results**
- **Discussion**
- Conclusion and future work
- References

What is a good paper?

- Problem is clearly described and interesting (technically hard, impactful, etc.)
- Definitions are clear and precise
- Original contribution is provided compared with related works
- Protocol is clearly described, reproducible, satisfying
- Experimental results that support the propositions
  - nicely described and analysed
- Related works are well cited and correctly discussed

- So, what is a bad paper?

Outline

- Why would you read a paper?
- What are papers?
- **How to read a paper?**
- Results and evaluation
- Writing while reading

So many papers and still 24 hours in a day

- Any research field is rich of many papers
  - several thousands
  - impossible to read them all
- Still, necessity to produce scientific knowledge
  - situated *wrt* the state of the art

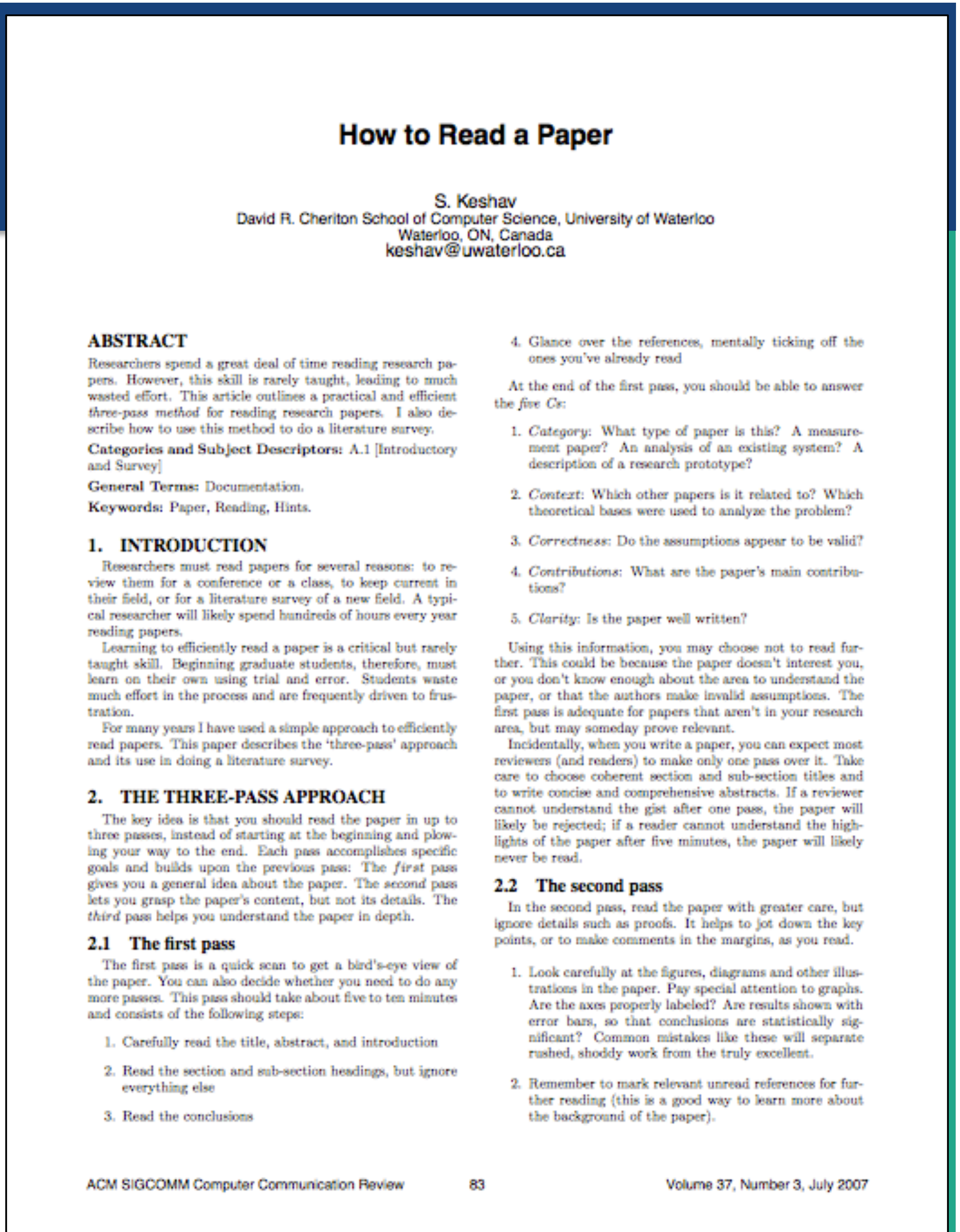
- Step 1: find articles (see dedicated course)
  - read title, abstract and keywords to decide if downloading the full article is needed
- Step 2: read the articles
  - but still many articles

Active reading

- Get rid of distractions
- Get a pen
- Jump around, re-read, go backward
- Talk to others

A three-pass approach

1. Getting a general idea about the paper **10 minutes**
2. Grasping the paper's content, but not its details **one hour**
3. Understanding the paper in depth **several hours**



How to Read a Paper

S. Keshav
David R. Cheriton School of Computer Science, University of Waterloo
Waterloo, ON, Canada
keshav@uwaterloo.ca

ABSTRACT

Researchers spend a great deal of time reading research papers. However, this skill is rarely taught, leading to much wasted effort. This article outlines a practical and efficient *three-pass method* for reading research papers. I also describe how to use this method to do a literature survey.

Categories and Subject Descriptors: A.1 [Introductory and Survey]

General Terms: Documentation.

Keywords: Paper, Reading, Hints.

1. INTRODUCTION

Researchers must read papers for several reasons: to review them for a conference or a class, to keep current in their field, or for a literature survey of a new field. A typical researcher will likely spend hundreds of hours every year reading papers.

Learning to efficiently read a paper is a critical but rarely taught skill. Beginning graduate students, therefore, must learn on their own using trial and error. Students waste much effort in the process and are frequently driven to frustration.

For many years I have used a simple approach to efficiently read papers. This paper describes the 'three-pass' approach and its use in doing a literature survey.

2. THE THREE-PASS APPROACH

The key idea is that you should read the paper in up to three passes, instead of starting at the beginning and plowing your way to the end. Each pass accomplishes specific goals and builds upon the previous pass: The *first* pass gives you a general idea about the paper. The *second* pass lets you grasp the paper's content, but not its details. The *third* pass helps you understand the paper in depth.

2.1 The first pass

The first pass is a quick scan to get a bird's-eye view of the paper. You can also decide whether you need to do any more passes. This pass should take about five to ten minutes and consists of the following steps:

1. Carefully read the title, abstract, and introduction
2. Read the section and sub-section headings, but ignore everything else
3. Read the conclusions
4. Glance over the references, mentally ticking off the ones you've already read

At the end of the first pass, you should be able to answer the *five Cs*:

1. *Category*: What type of paper is this? A measurement paper? An analysis of an existing system? A description of a research prototype?
2. *Context*: Which other papers is it related to? Which theoretical bases were used to analyze the problem?
3. *Correctness*: Do the assumptions appear to be valid?
4. *Contributions*: What are the paper's main contributions?
5. *Clarity*: Is the paper well written?

Using this information, you may choose not to read further. This could be because the paper doesn't interest you, or you don't know enough about the area to understand the paper, or that the authors make invalid assumptions. The first pass is adequate for papers that aren't in your research area, but may someday prove relevant.

Incidentally, when you write a paper, you can expect most reviewers (and readers) to make only one pass over it. Take care to choose coherent section and sub-section titles and to write concise and comprehensive abstracts. If a reviewer cannot understand the gist after one pass, the paper will likely be rejected; if a reader cannot understand the highlights of the paper after five minutes, the paper will likely never be read.

2.2 The second pass

In the second pass, read the paper with greater care, but ignore details such as proofs. It helps to jot down the key points, or to make comments in the margins, as you read.

1. Look carefully at the figures, diagrams and other illustrations in the paper. Pay special attention to graphs. Are the axes properly labeled? Are results shown with error bars, so that conclusions are statistically significant? Common mistakes like these will separate rushed, shoddy work from the truly excellent.
2. Remember to mark relevant unread references for further reading (this is a good way to learn more about the background of the paper).

S. Keshav. 2007. How to read a paper. *SIGCOMM Comput. Commun. Rev.* 37, 3 (July 2007), 83-84.

Questions to keep in mind

- What questions does the paper address?
- Is the problem relevant?
- What are the main conclusions of the paper?
- What evidence supports these conclusions?
- What it the quality of the evidence?
- Why are the conclusions important?

Pass 1: overview

- Title + authors
  - What is it about + Where does it come from?
- Abstract
  - What was done, what is the contribution?
- Medium
  - What is the audience?

- Introduction / conclusion
  - What are the context, problem, results?
- Sections/subsections headings (+ glance at formulas & figures)
  - What is the paper general structure?
- References
  - Is it a serious paper?

Pass 1: at the end

- Can I answer these questions?
  - Category
  - Context
  - Correctness
  - Contributions
  - Clarity

  } **5 Cs**
- Do I need to go further?

Pass 2: get the author message

- Identify key parts and articulation of the reasoning
- Identify key intuitions, limitations, relevance of the work

- Read carefully the text
- Focus on figures and their legends
  - do not spend time on demonstrations
- Mark down new terms, questions that arise
- Select references you may would like to consult

Pass 2: at the end

- Am I able to summarize the paper / justify the results to someone else?
- Do I need to go further?

Pass3: careful reading

- Go again from beginning to end...

  …and jump backwards
- Identify and challenge every assumption
- Read proofs
- Get help from related papers if it gets too complicated
- Come back to it later

Pass 3: at the end

- Do I fully understand the work?
- Can I remake it virtually?
- What are hidden assumptions?
- Where would I do things differently?

- A full-fledged review should go that far
- A paper read that far should go to the “reference pool” of one’s work and influence it
- Living with a paper
  - Research papers summarize months or years of work in a few pages → came back, re-read, question, re-discover

Potential problems

- You don't understand the context and positioning
  → read basic references in the domain to situate it
- You don't understand the contribution
  → if the paper has been accepted there should be one
  → look a the related works, confront to the other approaches
- You don't understand the technical details
  →do you have to?
  →if yes, identify where is the problem to go deeper / seek help

Summary: navigate between

skimming ⟷ **deep reading**

idea of the subject ⟷ **good understanding**

Pass 1 Pass 2 Pass 3

Beyond PDF

synapse (Campi et al., 2005; Yokosuka et al., 2005; DeMond et al., 2008; Yu et al., 2012). Thus a lone agonist pMHC bound to TCR leads to stable engagement of the resulting complex with the cytoskeleton. Single molecule intensity calibration of the number of ZAP70 recruited to the vicinity of each agonist pMHC indicates that TCR are triggered in a 1:1 stoichiometry with pMHC.

Associations of pMHC with TCR exhibited molecular binding dwell times with mean durations of 53.8 ± 12.2 s and 5.2 ± 0.2 s for AND and 5c.c7 TCRs, respectively. Individual dwell times are roughly exponentially distributed and are in general agreement with bulk solution measurements of pMHC:TCR kinetic off-rates for both TCRs (Corse et al., 2010; Huppa et al., 2010; Newell et al., 2011). However, dwell times measured from tracking experiments specifically correspond to spatial entrapment of pMHC with a TCR, or cluster of TCRs (Schamel and Alarcon, 2013), on the T cell surface. They do not necessarily correspond to individual molecular binding events with a single TCR. Indeed, recent studies (e.g., by FRET) have suggested that pMHC:TCR kinetic off-rates may be accelerated in living cells relative to in vitro measurements, possibly as a result of actively applied forces from the cytoskeleton (Huang et al., 2010; Huppa et al.,

CONTENT

FIGURES

REFERENCES

INFO

Citation

TCR-peptide-MHC interactions in situ show accelerated kinetics and increased affinity

JB Huppa, M Axmann, MA Mortelmaier, BF Lillemeier, EW Newell, M Brameshuber

Nature, 463: 963-7, 2010

DOI: http://dx.doi.org/10.1038/nature08746

Citation

Spatial and temporal dynamics of T cell receptor signaling with a photoactivatable agonist

M Huse, LO Klein, AT Girvin, JM Faraj, Q-J Li, MS Kuhns

Immunity, 27: 76-88, 2007

DOI: http://dx.doi.org/10.1016/j.immuni.2007.05.017

eLife LENS - http://elifesciences.org/elife-news/lens

Outline

- Why would you read a paper?
- What are papers?
- How to read a paper?
- **Results and evaluation**
- Writing while reading

Examples of contributions

- A new way to solve a problem
- A new problem
- The proof of a theorem
- A data model
- Indicators of the behaviour of a system
- An new implementation of a known algorithm
- An interaction principle
- A language
- A new concept to describe usages
- A collaborative system
- A methodology for designing systems
- A study of how users behave with a system

Variety of contributions → Variety of justifications

Types of justifications

- Formal proofs
  - Mathematics, logics
- Experimentations
  - Experimental: discovering facts, investigating, testing hypothesis
  - Validation: proving that a theory/idea/algorithm/setting is correct
- Argumentation
  - Logical argumentation
- …

Formal Proofs

- Prove the theoretical complexity of an algorithm
  - Linear? Exponential?
- Prove the completeness and correction of an algorithm
  - Does it produce all the results?
  - Does it produce the wanted results?

- Needs
  - Formalisation of the data and the algorithm
  - (use of lemma)

Experimentations (1)

- Check the calculation time for an algorithm
  - To check if it matches theoretical complexity
  - To observe its functioning (when too complex)

- Needs
  - Precise time measure
  - Carefully chosen, convincing trial data
  - Precise evaluation methodology: running trial, managing result data
  - Definition of indicators

Experimentations (2)

- Show that a system is more efficient to do a task than another

- Needs
  - Define independent and dependant variables
  - Make hypothesis
  - Define protocol
    - e.g. subjects, intra or inter-group
  - Statistically check null hypothesis
    - $p < .05$ meaning

Experimentations (3)

- Show the efficiency of an information retrieval system

- Needs
  - Define a corpus of documents
  - Define some “ground truth”
    - for that query those are the docs that match
  - Evaluate precision and recall
    - fraction of retrieved docs that are relevant
    - fraction of relevant docs that are retrieved

- Variations: clustering, learning, etc.

Experimentations (4)

- Check what happens with a new system
- Check parameter sensitivity for an algorithm on the quality of results

- Needs
  - Build the system
  - Observe how it behaves
    - E.g. results, memory usage
  - Draw empirical conclusions

Argumentation

- Validate a model: check whether an ontology is appropriated for what it was conceived for

- Needs
  - Produce characteristics that define appropriateness
  - Define model or meta-model
  - Check the result against the characteristics
    - e.g. capacity to describe a domain, simplicity, extensibility

Benchmarks

- General idea
  - define standard shared datasets against which various approaches can be compared
  - mutualisation of the costs of building benchmarks
- Domains
  - Learning, clustering, information retrieval, language recognition, graph matching, etc.
- Contests
  - Public learning sets
  - Test sets used for comparing
  - e.g. TREC (Text REtrieval Conference)

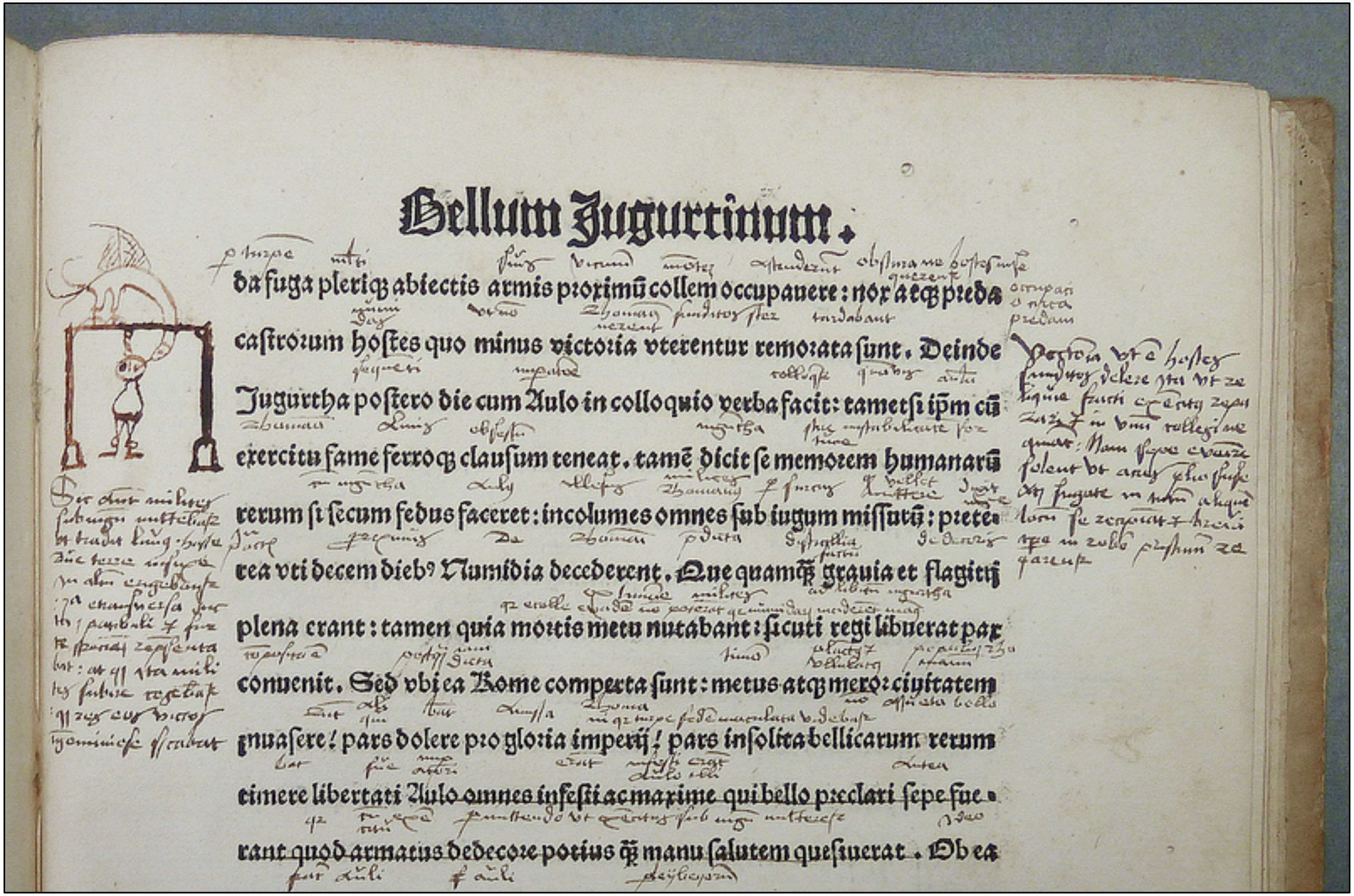
Outline

- Why would you read a paper?
- What are papers?
- How to read a paper?
- Results and evaluation
- **Writing while reading**

Active reading

- Not only sequential reading
  - skipping parts, getting back, see elsewhere
- Writing while reading
  - there is always a writing objective
    - possibly on the long term
    - possibly never to be written
  - annotating, taking notes is preparing the realization of that objective
- No standard method
  - Very idiosyncratic

Annotations



Annotations

classify (m) points represented by the matrix $A \in \mathbb{R}^{m \times n}$

2.2 Problem modelling

Our purpose is to classify m points of $\mathbb{R}^n$, represented by the matrix $A \in \mathbb{R}^{m\times n}$. Each point belongs to the class $+1$ or -1 depending on its classification in a diagonal matrix $D \in \mathbb{R}^{m\times m}$: the diagonal element $D_{ii} = +1$ if the point i belongs to the class $+1$, otherwise $D_{ii} = -1$ and the point i belongs to the class -1. For this problem, the standard SVM with a linear kernel is characterized by the following quadratic optimization problem for some parameter $\nu > 0$:

standard SVM with a lineal kernel

should be big

Is given: A matrix which contains the points to classify

variables: w, γ, y

$$\min_{(w,\gamma,y)\in\mathbb{R}^{n+1+m}} \quad \nu e^T y + \tfrac{1}{2}\|w\|_2^2$$
$$\text{subject to} \quad D(Aw - e\gamma) + y \geq e, \quad y \geq 0, \tag{2.1}$$

the variables y weighted by ν

Quadratic optimization problem

parameter $\nu > 0$

e is a vector of 1's of dimension m.

variables $y_i, i = 1, \ldots, m$ mean the error of classification of point

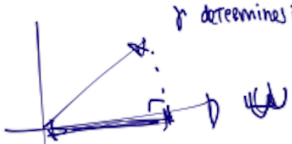
when both classes are strictly linearly separable then the values can be zero.

variables: w normal to the decision boundary; y means the error of the classification of the point; γ determines its location relative to the origin

where e is a vector of 1's of dimension m. (class, m points)

As shown in Figure 2.3, w is the normal to planes:

$$x^T w = \gamma + 1,$$
$$x^T w = \gamma - 1,$$



and γ determines its location relative to the origin. The plane on the top delimits the $+1$ class (stars) and the plane on the bottom delimits the -1 class (circles). Variables $y_i, i = 1, \ldots, m$ mean the error of classification of point i. When both classes are strictly linearly separable, then the values y can be zero. The separating surface is the plane

$$x^T w = \gamma,$$

$\therefore \; \gamma = \|X\| \cdot \|W\| \cos\theta = X_1 W_1 + X_2 W_2 + X_3 W_3 + \ldots + X_4 W_4$ (escalar product)

which is located in the middle of the two delimiting planes. If the classes are linearly non-separable (as in Figure 2.3), then both planes delimit the two classes with a "smooth margin" determined by the nonnegative variables y, so:

$$x^T w - \gamma + y_i \geq +1, \text{ for } x^T = A_{i.} \text{ and } D_{ii} = +1,$$
$$x^T w - \gamma - y_i \leq -1, \text{ for } x^T = A_{i.} \text{ and } D_{ii} = -1,$$

where $A_{i.}$ is the i-th row of A.

The objective function minimizes the one norm (or Manhattan distance) of the variables y

14 / 19

Active reading result (1)

Reading notes

- Goal
  - systematically keep track of the content of the papers that have been read
    - act as a proxy for the article
      - useful to refresh one's memory without re-reading the article
    - allow exchange about an article
      - with one's supervisor
      - within one's research team
- Difficulty
  - finding the good granularity
    - personal summary of article + important ideas for one's research
  - be systematic

Active reading result (1)
Reading notes

- Writing reading notes
  - Structure
    - Bibliographical information
    - Type of article
    - Subject and reason why it has been read
    - Main contributions
    - Summary: problem, solution, experimental setting, results, etc.
    - **Commentary**
      - good points, limits
      - lessons learned with regard to one's personal research interest
  - From 1 to 4 pages
    - Could be blogged!

Active reading result (2)
Synthesis / literature review

- Goal: build an exhaustive personal view of a scientific sub-field
  - First task of a master's thesis, a PhD
  - Will be used
    - in "Related work" sections of articles
    - in the first part of your thesis
    - (in a review article)
- Difficulty: finding the appropriate articles
  - Core articles and side articles
  - Triangulate: spot
    - articles cited by everybody
    - important authors (see webpages)
    - important conferences (skim through program)

Active reading result (2)

Synthesis / literature review

- Writing Synthesis
  - Organize articles along themes / sections
    - an article can be cited in several themes
    - draw conclusion at the end of each section
  - When it comes to describing systems
    - summarize in table with characteristics that help to compare, only describe exemplar systems
  - Terminate with remarks related to your research
    - Directions you should take

Active reading result (3)

Article review

- Goal
  - Give your opinion on the acceptance of an article for publication
    - journal / conference
  - Give advice on how to improve the article
  - Most of the time anonymous
- Difficulty
  - Takes time
  - Not easy
    - e.g. assessing interest of a paper vs assessing technical quality

Active reading result (3)
Article review

- Writing Article reviews
  - Necessitates deep reading (pass 3)
  - There are guidelines
  - Advice
    - Take your time
      - Read one day, re-read and write the review another day
    - Be careful of what you write (be gentle)
      - Would you accept to get that comment?
      - Some reviewers drop anonymity
    - Be ethical
      - Do not use original material you read while doing a review
      - Do not accept to review in case of conflict of interest

  - See http://www.scoop.it/t/toolsandmethodologyforresearch for advice