Georgia Tech

Lecture 08. Machine Learning Project

Xin Chen

Outline

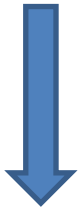
- A big picture: how to identify a problem?
- A good start: pay attention to your intro and background.
- General guidance for a project, including proposal, presentation and a report.
- How to choose topic for a machine learning project?
- Specific requirements for the proposal, presentation and the report.

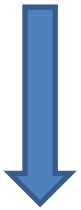
Outline

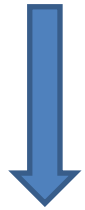
- A big picture: how to identify a problem? 
- A good start: pay attention to your intro and background.
- General guidance for a project, including proposal, presentation and a report.
- How to choose topic for a machine learning project?
- Specific requirements for the proposal, presentation and the report.

Identify a problem

- **Why?** (why is this problem interesting?)
  - Why did you do all this work?
  - Why does it need a solution?
  - Why do other people want to read it?
  - Why do other people care about it?
- **What?** (what can you contribute?)
  - What difference do you make on this problem?
  - What novel thing do you do?
  - What is your main idea?
  - What can others learn from it?
- **How?** (how does your approach work?)
  - How do you practically/theoretically implement your approach?
  - How do you deal with the problem? Challenge and novelty
  - How does your approach perform better than others?
  - How can others benefit from it?







A Bad Example

- Why the problem is an interesting problem?
  - Deep learning is everywhere, so it is an interesting problem.
- What can you contribute?
  - I read all the source code of Linux kernel or Tensor flow
  - I design an algorithm with 100K line of codes
- How does your approach work?
  - I use deep neural network, so it becomes better.

A Good Example: design a food

- Why the problem is an interesting problem?
  - Many people like healthy and fast food
  - However, most food are either unhealthy but fast, or healthy but slow.

- What can you contribute?
  - I want to “design” a food from scratch.
  - The key idea is to provide a fast, healthy and delicious food, by using some ingredients and some special cooking method.

- How does your approach work?
  - The ingredients of the new food is salt etc.
  - The steps of the cooking is A, B, C …

A good example in Machine Learning

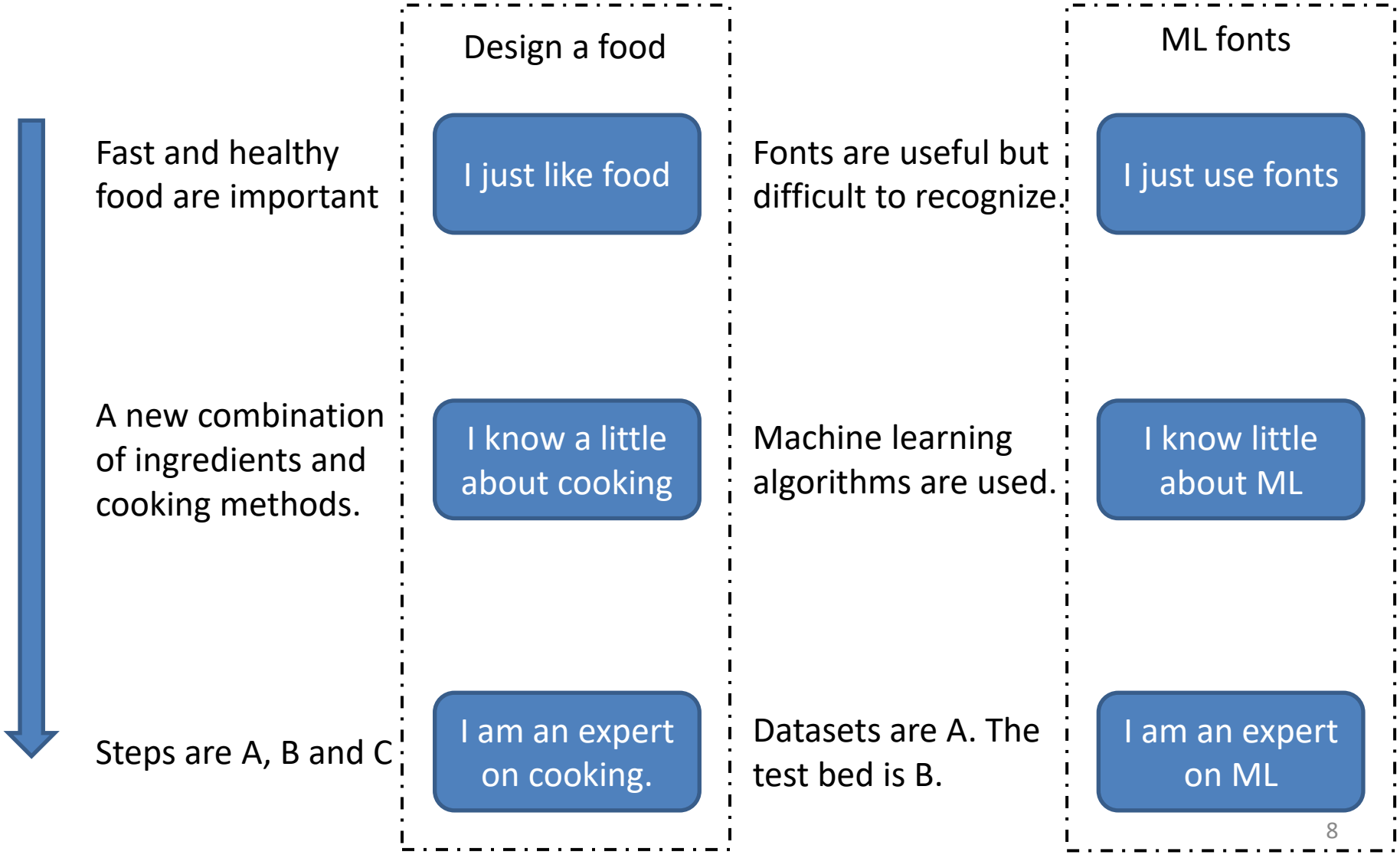
- Why the problem is an interesting problem?
  - Fonts make words more expressive.
  - Beautiful fonts cause inconvenience that possibly cause human errors.

- What can you contribute?
  - I want to convert images with 100,000 fonts to text.
  - The approach uses machine learning algorithms, including logistic regression, random forest and convolutional neural network (CNN).

- How does your approach work?
  - How do we set up the experiments?
  - Inputs: synthetic data, benchmark, baseline algorithm.

https://chrisgrizzly.github.io/font.teller/

Why does it look like this way



Outline

- A big picture: how to identify a problem?
- A good start: pay attention to your intro and background. 
- General guidance for a project, including proposal, presentation and a report.
- How to choose topic for a machine learning project?
- Specific requirements for the proposal, presentation and the report.

Place a good start in intro and background

- Intro and background are the first two sections that start you presentation or report.

- When thinking about your problem, you need to have a good start.
  - If you have a good start, everything else follows well.
  - If you have a bad start, it would become extremely difficult to bring it back on track, no matter how novel your idea is or how much effort you put there.

- Several key questions:
  - What is your problem that interests other people?
  - What is your new angle?
  - Have you described clearly how others solve this problem?
  - Comparison as much as possible between your approach and others'.

Intro and background is important

- Intro and background seems unimportant. Why?
  - Because we are “lazy”.
  - When you want to describe an object or a person to others with the shortest sentence, you intuitively use the most shinning point to represent the thing.
    - The superman who saved this city is …
    - The object that run very fast is …

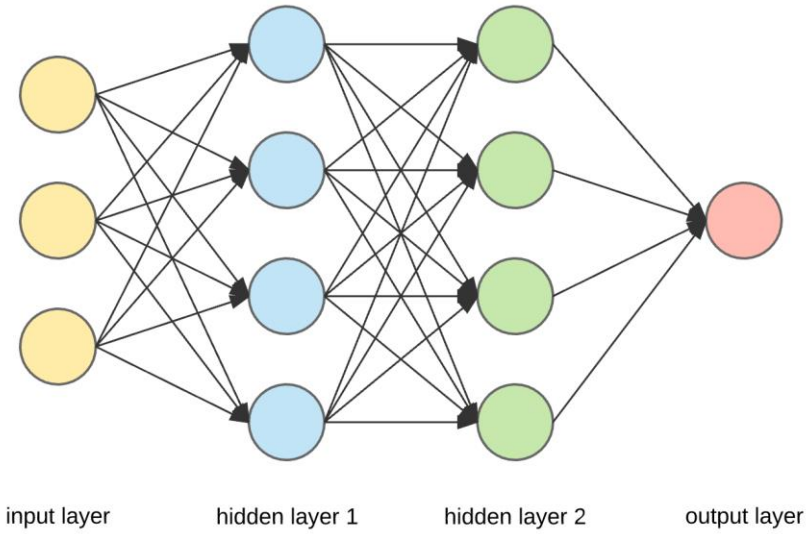
- We are a story-teller.
  - Intro and background is our first step.
  - A claim without a reasonable context is not convincing.

How to start?

- Intro and background
  - Assume that your readers and audiences have known certain information.
  - The starting point: this is the right way or the perspective to think about the problem I am solving.
  - This is the place that you can have numerous variances. For the same problem, you can have many ways to look at the problem as the “new angles”.

A Bad Example

- Neural network for image classification
- Related work
  - KDD paper 1 …
  - ICCV paper 2 …



- If you have known these papers, you still do not know what this problem is about.
- If you do not know these papers, nothing is told.

A good start

- A good start may look like:
  - Maybe you have known the technique “A” that can solve the problem “B”, but we found out that “A” can solve the problem “C”.
  - Or the problem can be solved by another technique “D”.

- The pattern is like:
  - We know something, but you may don’t know another thing, which a new angle you can look at our problem.

Outline

- A big picture: how to identify a problem?
- A good start: pay attention to your intro and background.
- General guidance for a project, including proposal, presentation and a report. 
- How to choose topic for a machine learning project?
- Specific requirements for the proposal, presentation and the report.

General guidance for a project

- Motivation of a talk/report
  - To tell an interesting story that you'd like to share. (even if you are forced to do so)
  - Collect high-quality feedback
  - Build more connection for collaboration, startup

- Clarity is always the highest priority.
  - Everyone is lazy and will be lazy, especially during a talk. People prefer not to think (much).
  - Being concise. Each slide plays a role.
  - Pay attention to your intro and background
  - Be clear about your problem: input, output, and constraint.
  - Establish goals and assumption early. Be consistent.

General guidance for a project

- Do not think readers/audiences are “smart”.
- Pay attention to details as much as possible.
  - 1 high-quality project >> 10 low-quality project
- Baseline is everywhere.
  - There is always a baseline when you say “good” or “bad”.
  - Be clear about your comparison. (what do you compare, how do you compare).
- Do not always expect an excellent result
  - You just need to figure out the reasons for a bad result.
  - A bad result is also very good, because you can avoid a bad path, which increases the probability of achieving a good result.

We all are “lazy” !!!

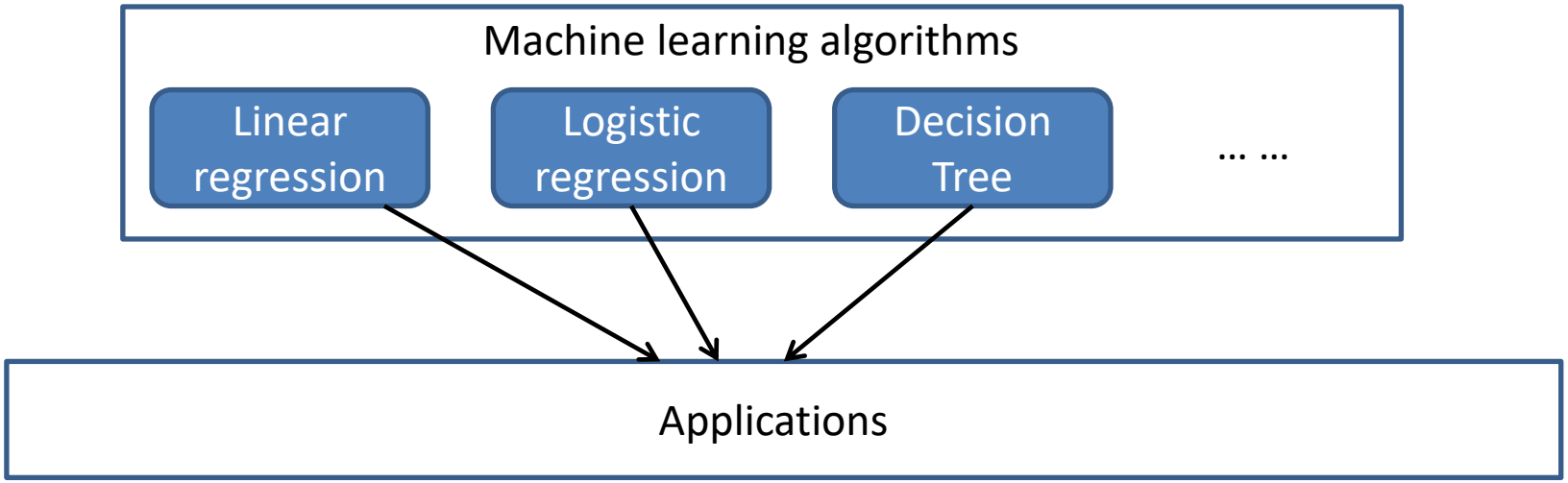
- Do not surprise me!
  - Arrange your story flow smoothly. (why, what and how)
  - Minimize difference between two slides or two paragraph.
  - Directly tell the readers, because we are too lazy to guess.
  - Tell the readers how to interpret your figures and plots. Do not think that it is easy to understand.

- Some surprise is pretty good ???
  - Our result is surprisingly good.
    - There is always an explanation for a good result.
    - You have a surprised result, because your context is well set.
  - Our proposed method is novel and we are the first that come up with this idea.
    - Everything have been reiterated for numerous times.
    - A novel idea always come from classic things. Your context should not surprise us.

Outline

- A big picture: how to identify a problem?
- A good start: pay attention to your intro and background.
- General guidance for a project, including proposal, presentation and a report.
- How to choose topic for a machine learning project? 
- Specific requirements for the proposal, presentation and the report.

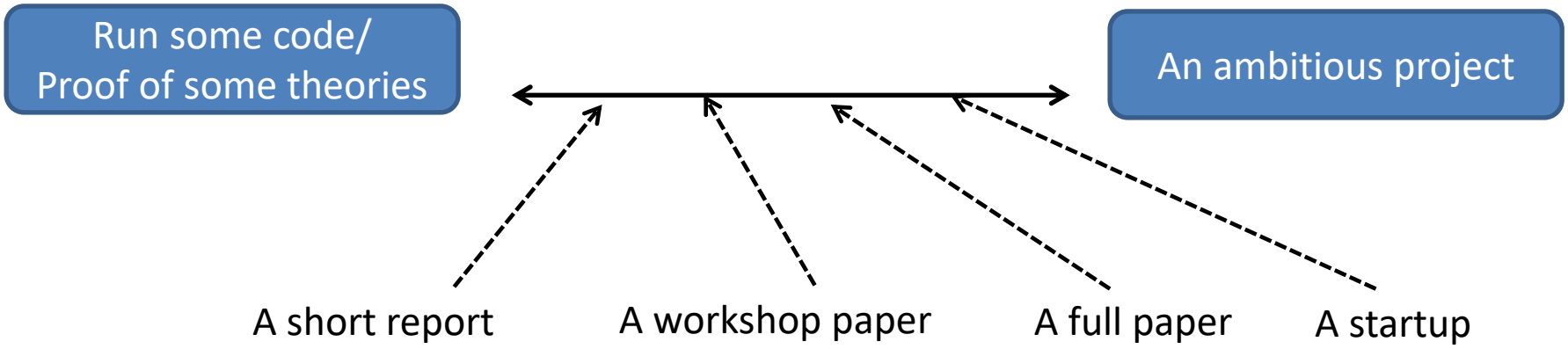
Machine learning algorithms



- Directions
  - Papers
  - Your interest (applications, hot topics)
  - Available datasets

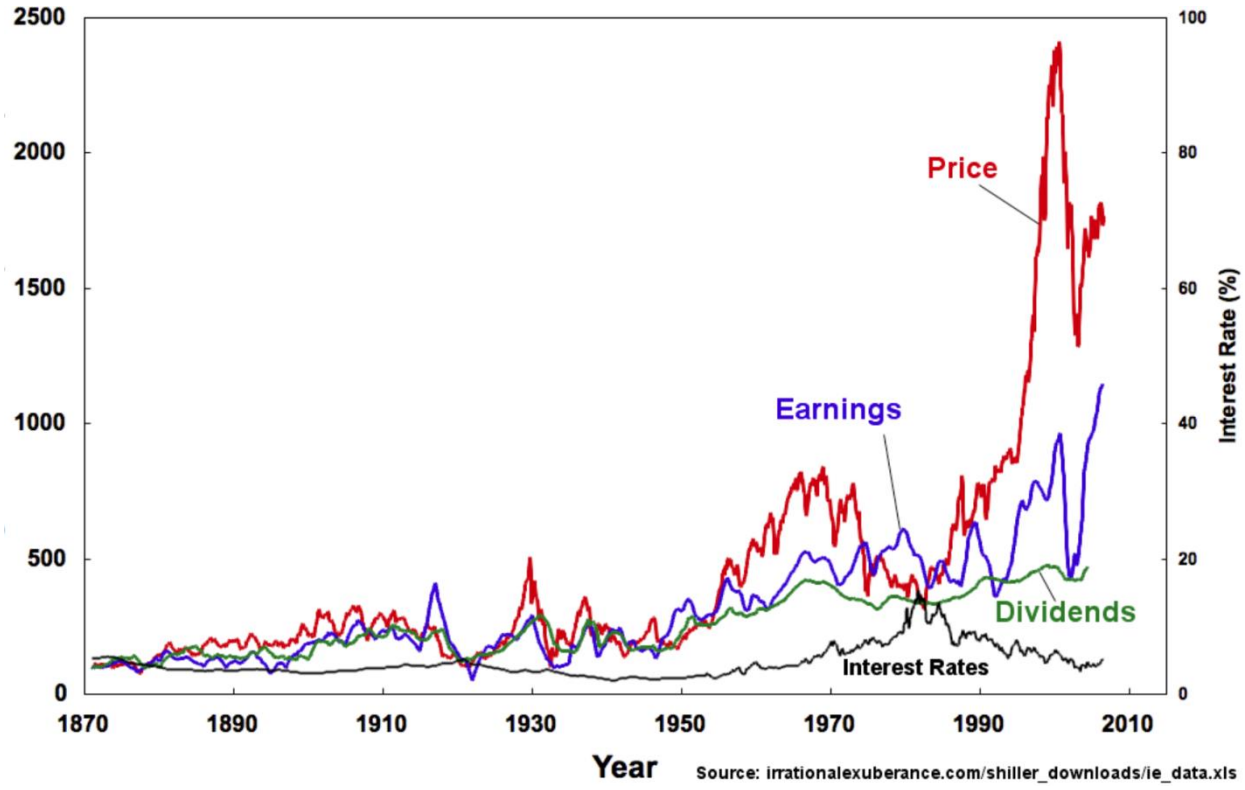
Please take this project seriously. This is your only chance in this course that you can explore in a deep way.

How to choose a ML topic



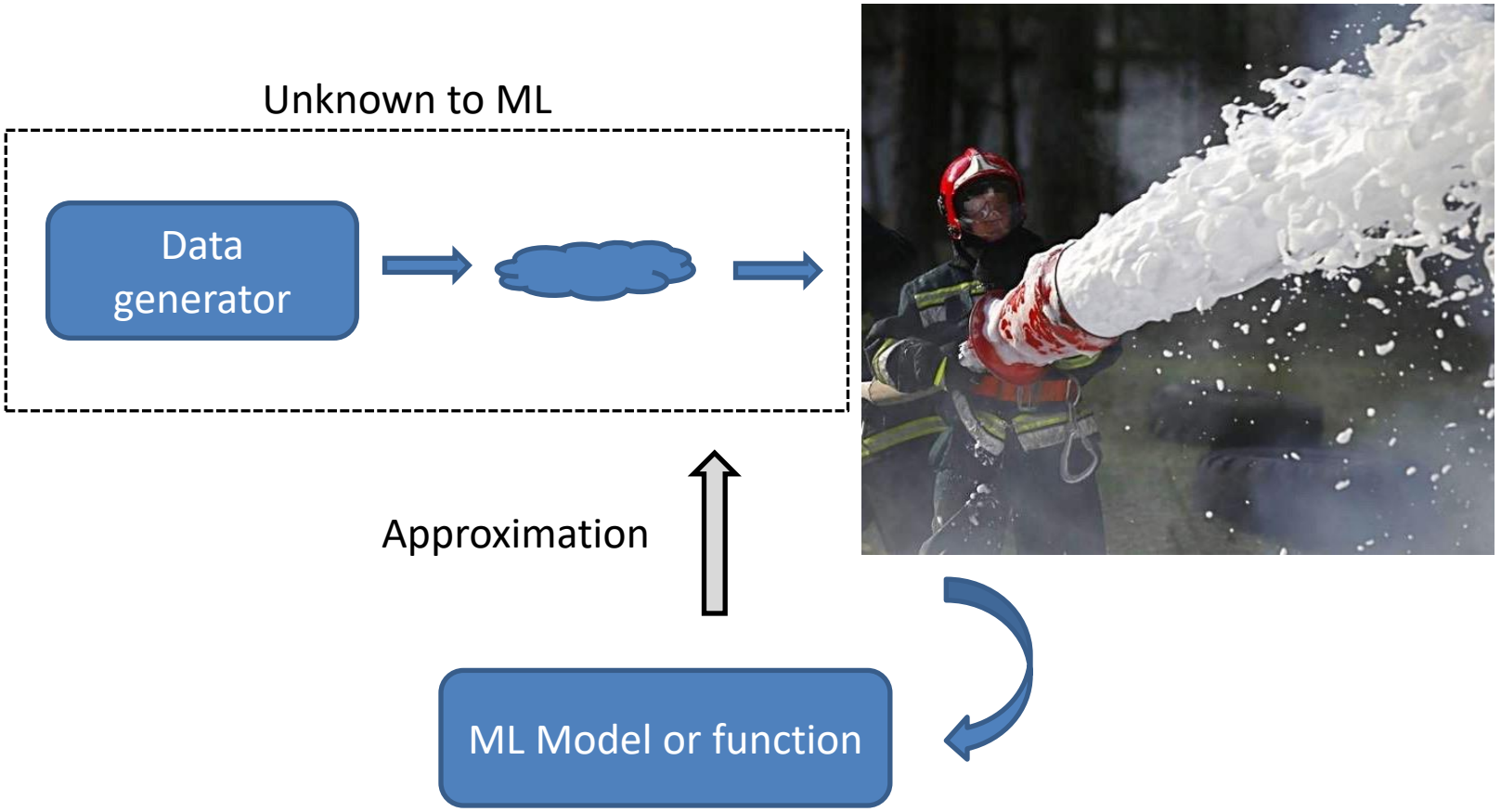
- Start from small and aim at big
  - It can focus on some ML algorithm.
  - It can focus on some dataset, or build a website
  - It can be a theoretical ML problem.
  - It can an implement of an algorithm from a paper.

Is predicting stock prices a good topic?



We have a large amount of datasets that can be used for training and test.

Is predicting stock prices a good topic?



Is predicting stock prices a good topic?

- My answer is No and Yes.

- No
  - There does not exist a rule or a principle that determines how to generate stock prices.
  - The operations of selling and buying stocks are all human behaviors.
  - An opposite example that can apply ML is to approximate some principle, like Newton's second law.

- Yes
  - If you think that human's behaviors always follow a pattern, your ML model is to approximate the pattern, then you can use ML to predict stock prices.

Outline

- A big picture: how to identify a problem?
- A good start: pay attention to your intro and background.
- General guidance for a project, including proposal, presentation and a report.
- How to choose topic for a machine learning project?
- Specific requirements for the proposal, presentation and the report. 

Project proposal

- Background (50% pts)
  - Define your problem.(by a definition or giving an example)
  - Why is it interesting to other people?
  - How do other handle it? (related work: including related problem, related paper, related techniques/algorithm)
- Objective (30% pts)
  - What do you want to do? (your focus)
  - List your potential contributions after you finish the project
- Plan (20% pts)
  - Timeline: when to do what
  - Responsibility: who to do what

Presentation

- There is no specific requirement on how to organize the presentation or the slides.

- Grading: (I don’t have a role here)
  - Each team needs to grade other teams’.
  - If you don’t grade other team, you will miss a portion.
  - The final grade is an average number, after removing the highest and the lowest.

- Please be prepared for the following questions:
  - What is your new angle to the problem?
  - How is your approach different from others?
  - Can you fully explain your results (good or bad)?
  - What is your learned lesson?

Some classic tips for a presentation

- One point per slide !!!
- Title matters.
  - It should be a great summary of the slide.
- Always, always, always explain your figure or graph.
  - How to interpret the figure should be told specifically.
- Build stages with section slides
- Audience is always right. Do not be defensive.

Report

- Introduction (15% pts)
  - State your problem
  - Why do people care or can possibly use your approach?
- Background (25% pts, including the references)
  - Related work on how other solve the problem
  - Compare these related work with pros vs cons
  - Baseline/Motivate your approach
- Design & results (50% pts)
  - Describe your approach on (Why, What, How) (25%)
  - For instance you design a new ML algorithm based on a classic algorithm (as a modification)
    - Why do you modify the algorithm in this way?
    - What do you add or remove?
    - How does it work? Especially how it work with other components?
  - Results: (25%)
    - Test bed information
    - The input and output, what dataset do you use?
    - How do you tune the parameters?
    - Your baseline algorithm as comparison
    - Plot your results
    - Explanation for your results
- Summary & conclusion (10% pts)
  - Summarize your approach and your result
  - What do you learn from this? Or what others can learn from this?
- References

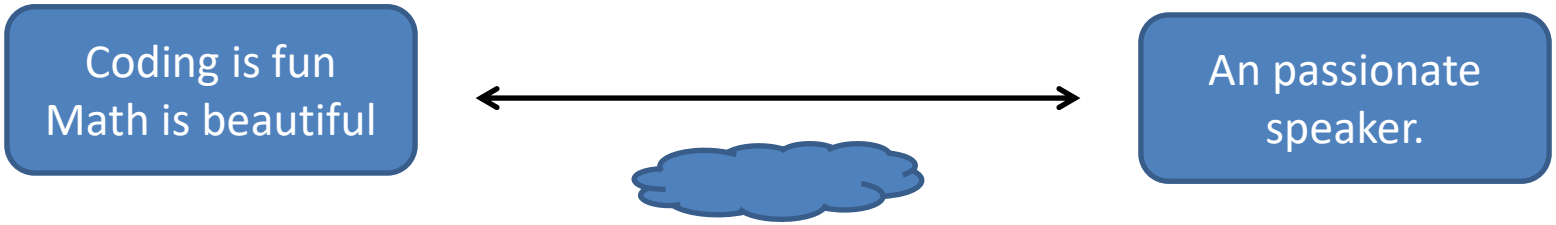
Why need to present and report?

- On a course, student needs to get a “A”.
- That’s it !!! ???

I think a lot and I like coding or math proof by myself.
I feel no need to present or report anything to others.

I am highly confident with my communication skills. I feel no need to understand the structure of the algorithms or techniques.

Why need to present and report?



There is a middle point between them, which is to present and report

Obviously, present and report can help you move from one end to the other.

More importantly, it helps you organize your knowledge in a clear way so that you are on a good path.

Take-home message

- How to identify a problem.
- How to organize a presentation and a report
- Hope you can learn something useful through this project, not only on the course but also you future career.
- Next:
  - I will assign you a team randomly on Friday **Jun 5th**.
  - Project proposal is due on Sun **Jun 14th.**
  - Project presentation is on Wed **July 13th**.