MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE

How to start your academic research life: strategy and practice

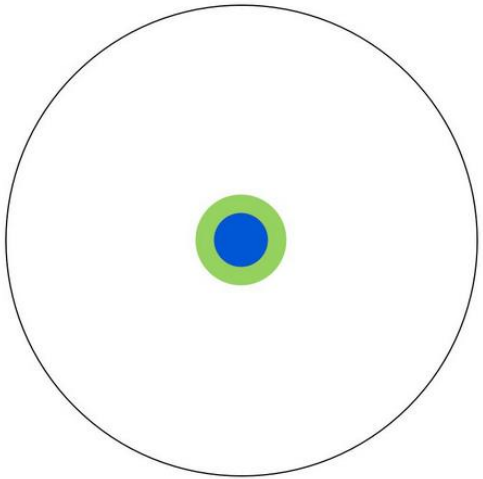
Bin Gu

Outline

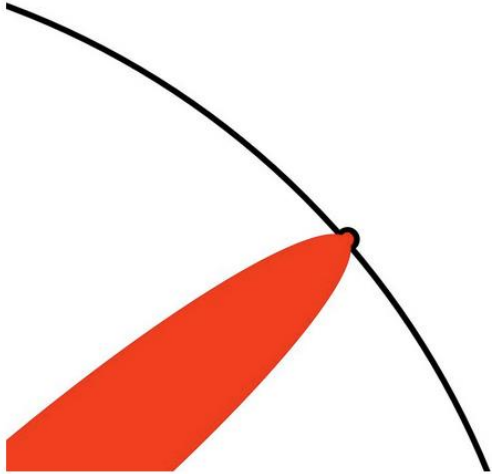
- What is academic research
  - Different learning /thinking styles in undergraduate and postgraduate studies
- My personal research journey
- Our strategy: learn through research
- My experience of supervising students
- Glorious future

What is academic research

Imagine a circle that contains all of human knowledge:



With a bachelor's degree, you gain a specialty:



And, that dent you've made is called a Ph.D.:

The boundary of human knowledge was pushed a little bit due to your unique research works

Research is like fish hawk fish: two stages

Circling

shutterstock.com · 1014679363

Catching

shutterstock.com · 155698367

Differences between undergraduate and postgraduate studies

- Outcome
  - Undergraduate: Grades, GPA, reward, competition
  - Postgraduate: Research paper, patent, software package…
- Problems
  - Undergraduate: designed by lecturer ----Closed world
  - Postgraduate: found by you or your collaborator ---- open world
- Practice
  - Undergraduate: Assignment, course project, exam--- Lightweight
  - Postgraduate: research ---- Heavyweight

Thinking style

- Thinking style
  - Critical thinking
    - Undergraduate: Not so much
    - Postgraduate: strongly dependent

  - Creative thinking
    - Undergraduate: Not so much
    - Postgraduate: strongly dependent

Period

- From having a problem or idea to having a good/full solution, and finally have a paper, normally it lasts several months, or years

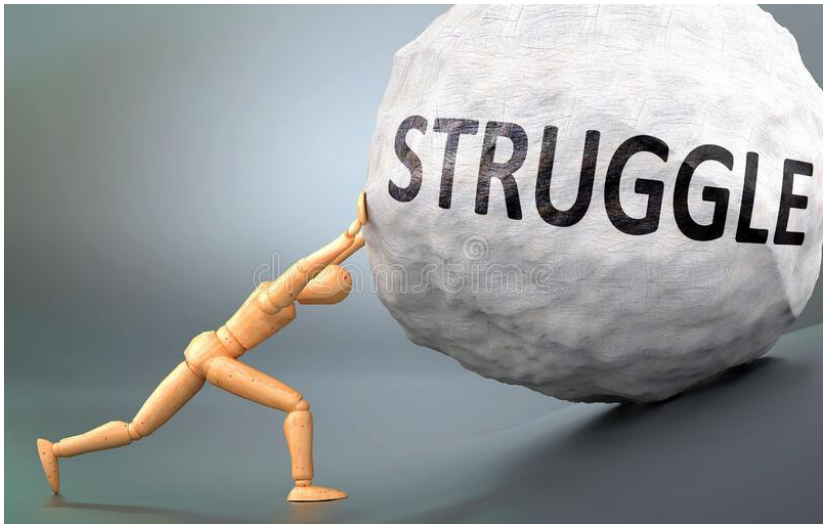
- This is a systematical work, we should make a plan. Just like building a house

Start your research is always not easy…

- First step:
  - Balance your time between course study and academic research

My personal research journey

- 2005 – end of 2014

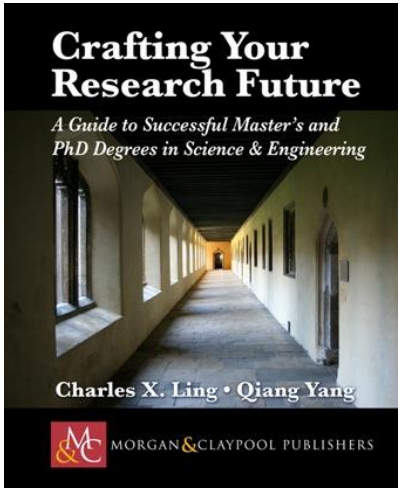


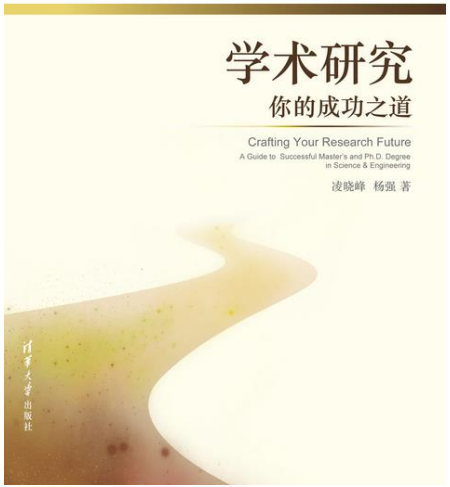
- 2014 – Now

Key turning point- end of year 2014

- Visiting research with Charles Ling for two months

Charles Ling





A New Generalized Error Path Algorithm for Model Selection

Abstract

Model selection with cross validation (CV) is very popular in machine learning. However, CV with grid and other common search strategies cannot guarantee to find the model with minimum CV error, which is often the ultimate goal of model selection. Recently, various solution path algorithms have been proposed for several important learning algorithms including support vector classification, Lasso, and so on. However, they still do not guarantee to find the model with minimum CV error. In this paper, we first prove that a large class of error (or loss) functions are piecewise constant, linear, or quadratic w.r.t. the regularization parameter, based on the solution path. We then propose a generalized error path algorithm (GEP), and prove that it will find the model with minimum CV error for the entire range of the regularization parameter. The experimental results on a variety of datasets not only confirm our theoretical findings, but also show that the best model with our GEP has better generalization error on the test data, compared to the grid search, manual search, and random search.

1. Introduction

In machine learning, most of learning algorithms are parameterized. For example, support vector machines (SVM-) (Vapnik, 1998) have a regularization parameter controlling the trade-off between a large margin and a small error penalty. Lasso (Tibshirani, 1996) has a regularization parameter on model's L_1 penalty to lead to sparse solutions. It is obvious that one fundamental task of the parameterized learning algorithms is model selection: tuning the parameters of models to achieve optimal generalization performance.

Model selection with cross validation (CV) (Arlot et al., 2010) is very popular in machine learning. CV is to split data, once or several times. For each split, one part of data (the training samples) is used for training the algorithm, and the remaining part (the validation samples) is used for estimating the error of the obtained model. The CV errors are averaged over the rounds. Then, CV selects the parameter among a group of candidates with the smallest CV error. Because of the simplicity and the universality, CV is a widespread strategy for model selection (Arlot et al., 2010). For example, Foster et al. (2014) used leave-one-out CV on SVM for medical diagnosis. Jahrer & Töscher (2012) used 16-fold CV for collaborative filtering. Usai et al. (2009) used repeated random sub-sampling validation on Lasso for genomic selection.

As mentioned above, CV works with a group of candidates of parameter. Normally, the candidates of parameters are specified by some strategies. The most popular strategy is grid searching. For example, the regularization parameter C of SVM is searched on a 18 grid linearly spaced in the region $\{\lfloor \log_2 C \rfloor - 9 \leq \log_2 C \leq 8\}$, as used in Yang & Ong (2011). Grid search is reliable in low dimensional parameter spaces. For high dimensional parameter spaces (such as parameters in Deep Belief Networks (Hinton et al., 2006)), manual search (Hinton, 2010), and random search (Bergstra & Bengio, 2012) were used for CV. However, as we know, CV with grid, manual, and random search strategies only considers finite candidates due to the limited computing resources. It can not guarantee to find the model with the minimum CV error in the whole parameter space, which is often the ultimate goal of model selection.

In this decade, a novel learning methodology called solution path (Hastie et al., 2004) has been developed for tracing the solutions with respect to a parameter of the parameterized learning algorithms. Fig. 1 shows the solution path of Lasso with respect to its regularization parameter. As shown in Fig. 1, one solution can act on an interval of the regularization parameter in which the solutions share a same linearity property. Thus, solution path algorithm can effectively give the entire solutions based on a finite number of representative solutions (the solutions at knee points). Solution path is totally different to the grid, and other common search strategies in which one solution only acts on one value of the parameter. Solution path algorithm has been proposed for several important learning algorithms. For example, Hastie et al. (2004); Gunter & Zhu (2007); Wang et al. (2008); Gu et al. (2012) proposed solution path algorithms for different types of SVMs. Rosset & Zhu (2007) proposed solution path algorithm for Lasso,

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

PAGE 1 10

c Dec 18
candidate parameters
[change the rest of the paper]

c Dec 18
which are continuous.
[needed? if too obvious, no need]

c Dec 18
searched?

c Dec 18
new

c Dec 18
often

c Dec 18
as?

c Dec 18
represent ... based on a finite set of numbers (such as the "knee points").
[??... "knee points" common term?]

c Dec 18
much more general as it represents an invite space, than ...
?? [more accurate statement is better]

c Dec 18
informal use

c Dec 18
new paragraph?
The solution path algorithms

PAGE 2 2

c Dec 18
as they do not correspond to "error path" in CV. That is, previous solution path algorithms do not lead to global minimal CV error.

Why the turning point works on me

- Take 9 years to prepare myself
  - In the initial 9 years, I worked hard even no many papers were published
  - I have done some interesting works, even most of them were published 3-5 years later.
  - Keywords: Never give up, work hard always
- Some support/guidance from your supervisor
  - The experience conducted from your supervisor
    - Congratulation: You become better!
  - If not work hard, the experience/guidance from supervisor does not work.

2016-2018

- UTA and Pittsburgh.
- Supervisor: Prof. Heng Huang
- Supervisor suggest me to work on the optimization
  - Challenges- new area to me
  - Work hard
  - 12am, 1am, 2 am, 3am, 4am, 5am, 6am in Pittsburgh and Arlington
- After I finish my second Postdoc, I become much better.

Asynchronous Doubly Stochastic Group Regularized Learning
First paper in stochastic optimization

$$\min_{x\in\mathbb{R}^n} F(x) = f(x) + g(x) \qquad (1)$$

where $f(x) = \frac{1}{l}\sum_{i=1}^{l} f_i(x)$, $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ is a smooth, possibly non-convex function. $g : \mathbb{R}^n \mapsto \mathbb{R} \cup \{\infty\}$ is

Algorithm 1 Asynchronous Stochastic Coordinate Descent Algorithm $x_J = \text{AsySPCD}(x_0, \gamma, J)$

Require: x_0, γ, and J
Ensure: x_J

1: Initialize $j \leftarrow 0$;
2: **while** $j < J$ **do**
3: Choose $i(j)$ from $\{1, 2, \ldots, n\}$ with equal probability;
4: $x_{j+1} \leftarrow \mathcal{P}_{i(j),\frac{\gamma}{L_{\max}} g_{i(j)}}\left(x_j - \frac{\gamma}{L_{\max}} e_{i(j)} \nabla_{i(j)} f(\hat{x}_j)\right)$;
5: $j \leftarrow j + 1$;
6: **end while**

Ji Liu and Stephen J Wright. 2015. Asynchronous stochastic coordinate descent: Parallelism and con vergence properties. *SIAM Journal on Optimization* 25, 1 (2015), 351–376.

Algorithm 1 Asynchronous Doubly Stochastic Proximal Gradient Algorithm with Variance Reduction (AsyDSPG+)

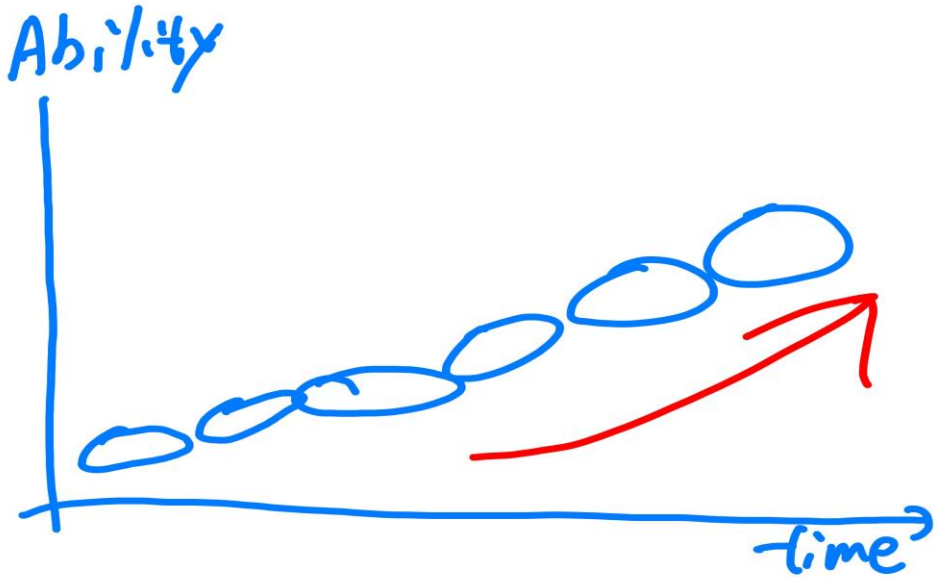
Input: The number of outer loop iterations S, the number of inner loop iterations m, and learning rate γ.
Output: x^S.

1: Initialize $x^0 \in \mathbb{R}^d$, p threads.
2: **for** $s = 0, 1, 2, \cdots, S-1$ **do**
3: $\widetilde{x}^s \leftarrow x^s$
4: *All threads parallelly* compute the full gradient $\nabla f(\widetilde{x}^s) = \frac{1}{l}\sum_i^l \nabla f_i(\widetilde{x}^s)$
5: *For each thread*, do:
6: **for** $t = 0, 1, 2, \cdots, m-1$ **do**
7: Randomly sample a mini-batch $\mathcal{B}$ from $\{1, ..., l\}$ with equal probability.
8: Randomly choose a block $j(t)$ from $\{1, ..., k\}$ with equal probability.
9: Compute $\widehat{v}^{s+1}_{\mathcal{G}_{j(t)}} \leftarrow \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} \nabla_{\mathcal{G}_j} f_i(\widehat{x}_t^{s+1}) - \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}} \nabla_{\mathcal{G}_j} f_i(\widetilde{x}^s) + \nabla_{\mathcal{G}_{j(t)}} f(\widetilde{x}^s)$.
10: $x_{t+1}^{s+1} \leftarrow \mathcal{P}_{\mathcal{G}_{j(t)},\frac{\gamma}{L_{\max}} g_{j(t)}}((x_t^{s+1})_{\mathcal{G}_{j(t)}} - \frac{\gamma}{L_{\max}} \widehat{v}^{s+1}_{t,\mathcal{G}_{j(t)}})$.
11: **end for**
12: $x^{s+1} \leftarrow x_m^{s+1}$
13: **end for**

Gu, B., Huo, Z. and Huang, H., 2018, March. Asynchronous doubly stochastic group regularized learning. In *International Conference on Artificial Intelligence and Statistics* (pp. 1791-1800). PMLR.

Strategy: learn through research

Multiple Closed cycle



One cycle

- Carefully set up your first research problem, not too hard, not too easy, according to your ability at that time
  - Do not hesitate to dig in your first project. You lose nothing.
- Do it
- If stuck the current work for 2-3 weeks, please discuss with your supervisor, if you should change a new problem
- Try to finish your work in two, four, or six months.

- --submit
- --revise/rebuttal
- --publish

Congratulation

Benefits of learning through research

- Deep understanding of one topic
  - You have deep understanding for multiple topics if you finish multiple closed cycle
  - Help you understand other topics deeply
  - Your knowledge on one topic is NOT isolated. It would help you to have a big picture on one topic some multiple topics
- You have some achievements (e.g. paper), after learning some materia

Additional discussion for collaboration

- Effective collaboration is not easy
  - Sense of team cooperation
  - you personally should be mature of researching

Ability + consciousness

Law of 10,000-hour

humans get really, really good at stuff through hard work.

My experience of supervising students with different levels

- Students from USA
  - Top students from Tsinghua, Peking … universities
    - Work hard + talent
    - Talent, but not work hard
- Students from middle level university
  - Xidian
    - Work hard
- Students from low ranked university
  - NUIST
    - Work hard

Conclusion: if you work hard and follow our strategy, you will be the best of you no mater what level of your original background.

Humbition

- Humble
- Ambition

It will help you to get help from senior researchers

Glorious future

- Please practice
- If having concern for our strategy, please do not hesitate to discuss with me.

- Hope to see your glorious future

Thanks for your attention!