Learning **for** Big Data

Hsuan-Tien Lin (林軒田)
htlin@csie.ntu.edu.tw

Chief Data Scientist @ Appier
Associate Professor @ CSIE of National Taiwan University

some sampled pages of my keynote talk
in IEEE BigData 2015 Taipei Satellite Session

About the Title

- “Learning **for** Big Data”

About the Title

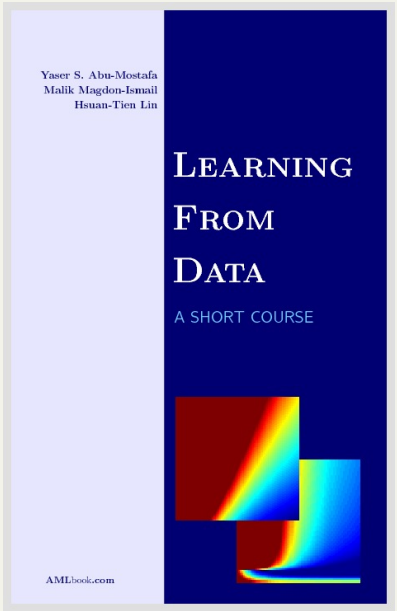
- "Learning **for** Big Data"
  —my wife: you have made a **typo**

About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?

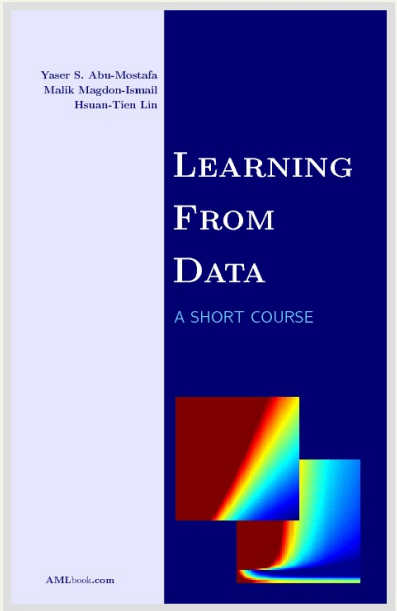
About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?
  —no, not a **shameless sales campaign**
  for my co-authored **best-selling** book ☺
  (`http://amlbook.com`)



About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?
  —no, not a **shameless sales campaign** for my co-authored **best-selling** book ☺
  (`http://amlbook.com`)

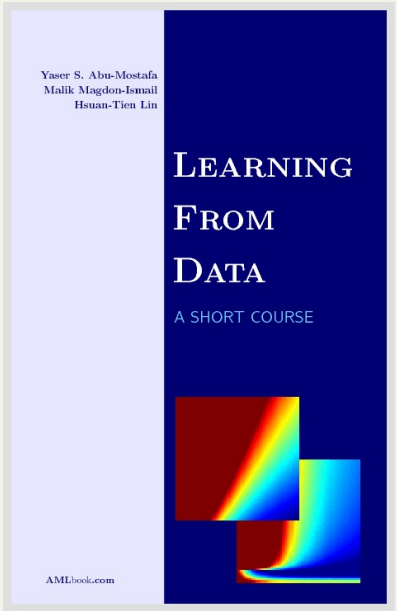


as machine learning researcher

machine learning **for** big data

About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?
  —no, not a **shameless sales campaign** for my co-authored **best-selling** book ☺
  (`http://amlbook.com`)

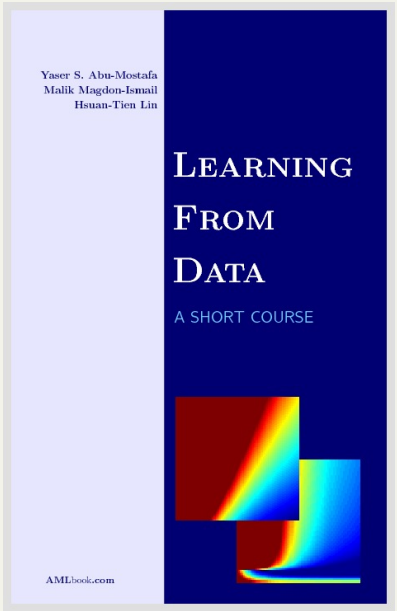


as machine learning researcher

machine learning **for** big data
—easy?! ☺

About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?
  —no, not a **shameless sales campaign** for my co-authored **best-selling** book ☺
  (`http://amlbook.com`)



as machine learning researcher

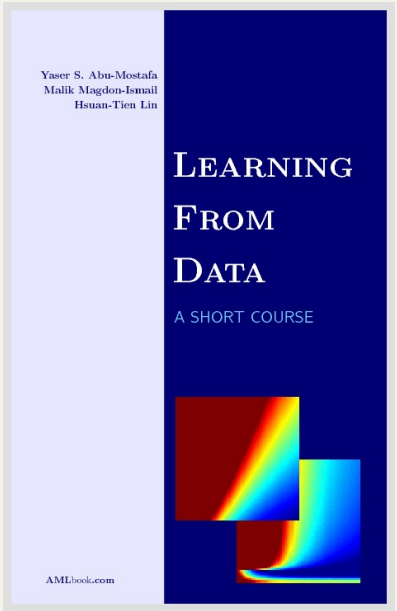
machine learning **for** big data
—easy?! ☺

as machine learning educator

human learning **for** big data
—**hard!!**

About the Title

- "Learning **for** Big Data"
  —my wife: you have made a **typo**
- do you mean "Learning **from** ~~Big~~ Data"?
  —no, not a **shameless sales campaign** for my co-authored **best-selling** book ☺
  (`http://amlbook.com`)



as machine learning researcher

machine learning **for** big data
—easy?! ☺

as machine learning educator

human learning **for** big data
—**hard!!**

will focus on **human** learning **for** big data

Human Learning for Big Data

Todo

- some FAQs that I have encountered as

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (), a AI-based startup

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (Appier), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

First Honest Claims

- must-learn for **big data** $\approx$ must-learn for **small data** in ML,

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

First Honest Claims

- must-learn for **big data** $\approx$ must-learn for **small data** in ML, but the former with **bigger seriousness**

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

First Honest Claims

- must-learn for **big data** $\approx$ must-learn for **small data** in ML, but the former with **bigger seriousness**
- system design/architecture **very important**,

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

First Honest Claims

- must-learn for **big data** $\approx$ must-learn for **small data** in ML, but the former with **bigger seriousness**
- system design/architecture **very important**, but somewhat beyond my pay grade

Human Learning for Big Data

Todo

- some FAQs that I have encountered as
  - **educator** (NTU and NTU@Coursera)
  - **team mentor** (KDDCups, TSMC Big Data competition, etc.)
  - **researcher** (CLLab@NTU)
  - **data scientist** (Appier), a AI-based startup
- my imperfect yet **honest** answers that hint **what shall be learned**

First Honest Claims

- must-learn for **big data** $\approx$ must-learn for **small data** in ML, but the former with **bigger seriousness**
- system design/architecture **very important**, but somewhat beyond my pay grade

I wish I had an answer to that
because I'm tired of answering that question.
—Yogi Berra (Athlete) ☺

Appendix: ML Foundations on NTU@Coursera

https://www.coursera.org/course/ntumlone

When can machines learn?

- L1: the learning problem (☺)
- L2: learning to answer yes/no (☺)
- L3: types of learning (☺)
- L4: feasibility of learning

Why can machines learn?

- L5: training versus testing
- L6: theory of generalization
- L7: the VC dimension (☺)
- L8: noise and error

How can machines learn?

- L9: linear regression (☺)
- L10: logistic regression (☺)
- L11: linear models for classification (☺)
- L12: nonlinear transformation (☺)

How can machines learn better?

- L13: hazard of overfitting (☺)
- L14: regularization (☺)
- L15: validation (☺)
- L16: three learning principles (☺)

☺ $\approx$ must-learn