Grading skills using machine learning

Shashank Srikant

shashank.srikant@aspiringminds.com

Aspiring Minds
research.aspiringminds.com

Research at Aspiring Minds

- Define product vision. Research and develop prototypes which demonstrate application of state of the art computer science technology in assessment products
- Publish at the very best conferences and venues
- Multiple outreach programs to popularize data science and machine learning
  - Data science for kids
  - ML-India
  - ASSESS – Annual workshop on data science for assessments
  - AMEO 2015

research.aspiringminds.com

CoDS 2016
Release AMEO-2016, public dataset on employability outcomes, based on AMCAT data

ACL 2015
Work on machine learning and crowdsourcing in speech evaluation

ICML 2015
Work on learning models for job selection

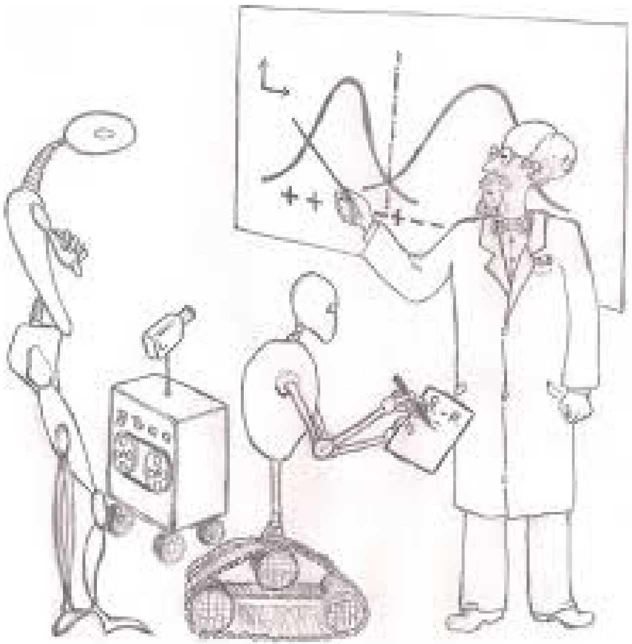
KDD 2014
Work on using machine learning in programming evaluation

NIPS 2013
Framework for using machine learning in assessments

aspiringminds
Employability Quantified

Machine learning

- Has caught the imagination of the public
- Important to understand intuition – handy tool
- The core setup is 12^{th} grade math!



ML for skill identification – Basic setup

Input → skills

↓

ML magic!

↓

Prediction → Ground truth

ML for skill identification – Basic setup

Input → skills

↓

ML magic!

↓

Prediction → Ground truth

ML for skill identification – Basic setup

Input skills

Language: Written & Spoken

Cognitive Skills

Functional Skills

Personality

Practical intelligence/Soft skills

ML for skill identification – Basic setup

Input skills

- Language: Written & Spoken
- Cognitive Skills
- Functional Skills
- Personality
- Practical intelligence/Soft skills

ML for skill identification – Basic setup

Ground truth

- Rubric
  - Rules provided by experts

Key learning 1

Thou shall suffer in setting up a gold standard

ML for skill identification – Basic setup

Input

Skills (quantified)

ML magic!

Prediction → Ground truth

Metric?

Type1 and Type2 errors – Metrics

Prediction → Ground truth

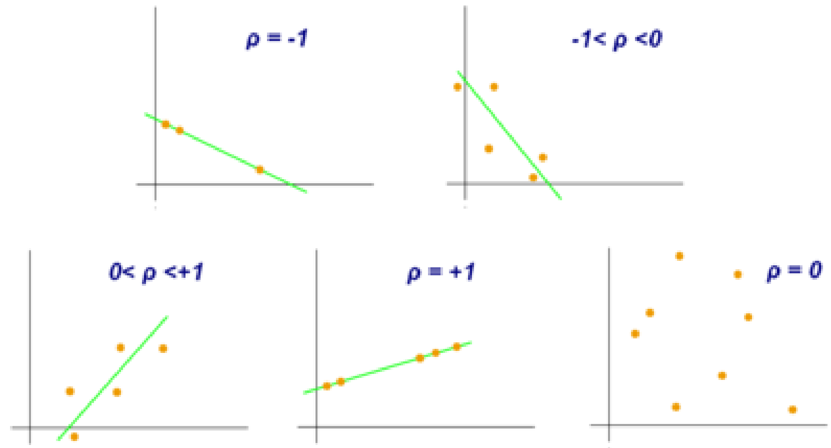
Metric?

		Reality	
		True	False
Measured / Perceived	True	Correct ☺	Type I False Positive
	False	Type II False Negative	Correct ☺

Pearson correlation – A metric

Prediction → Ground truth

Metric?



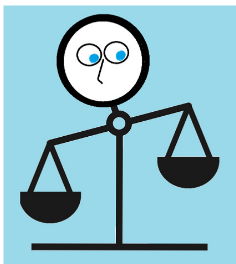
Pearson correlation – A metric

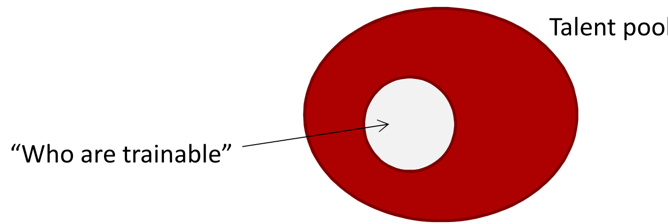
$$\underbrace{\text{Prediction} \longrightarrow \text{Ground truth}}_{\text{Metric?}}$$

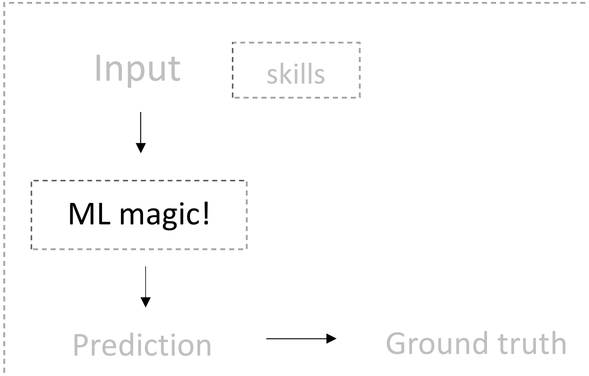
Trivia

Key learning 2

There's a perspective to skill grading when dealing with generic skill measures

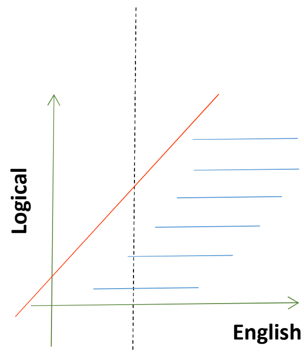


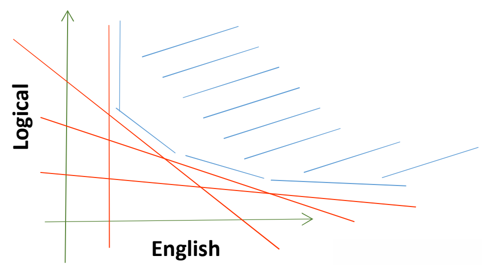


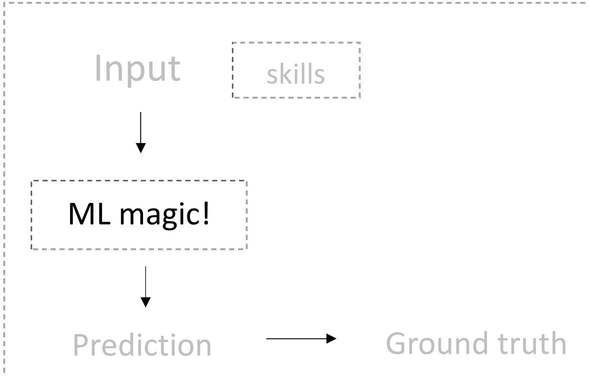


- Models need to be **interpretable**
- Should be **theoretically plausible**
- Models need to be **simple**
- Trade-off models between type 1 and type 2 error required

L > 400
E + 0.4*L > 870
E + 0.7*Q > 980





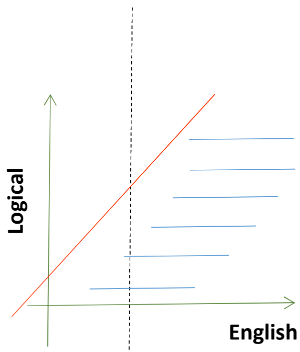


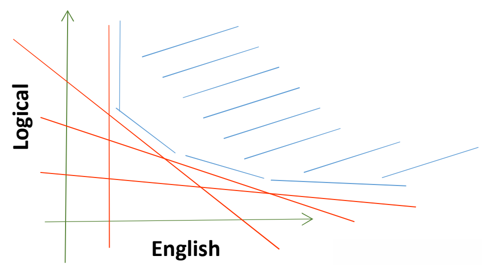
- Models need to be **interpretable**
- Should be **theoretically plausible**
- Models need to be **simple**
- Trade-off models between type 1 and type 2 error required

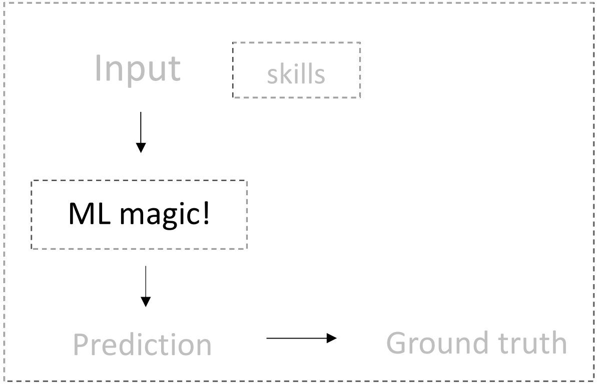
L > 400

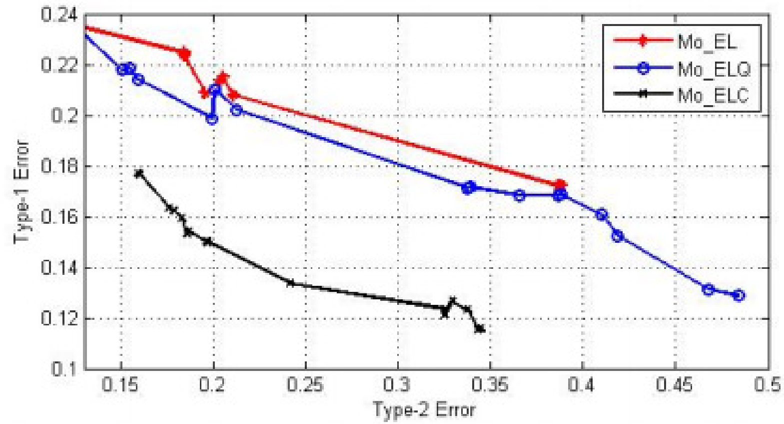
E + 0.4*L > 870

E + 0.7*Q > 980



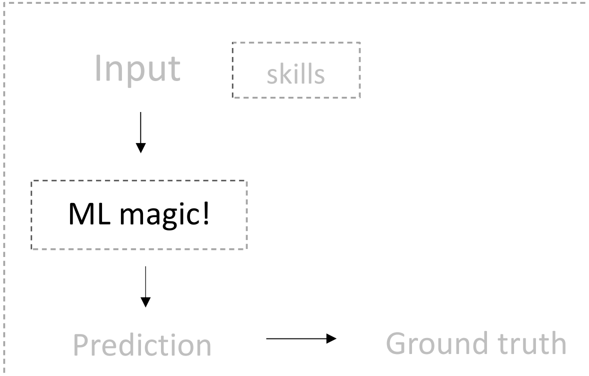






Type-1 vs. Type-2 between different models

aspiringminds
Employability Quantified



Model	Equation	Type-2 Error	Type-1 Error
PSO(E,L)	$L > 494$, $0.42 * E + L > 722$	0.21	0.21
PSO(E,L,Q)	$E > 405$, $L > 401$, $0.75 * Q + L > 874$	0.21	0.20
PSO(E,L,CP)	$CP > 360$, $0.47 * E + 0.98 * CP + L > 1196$	0.20	0.15

E : English comprehension, Q : Quantitative ability, L: Logical ability, CP : Computer programming ability

Key learning 3

Quant is seemingly useless in general cognitive tests

Thou shall suffer in setting up a gold standard

There's a perspective to skill grading when dealing with generic skill measures

---

Rubrics; Type1-Type2; Correlations; Interpretable models

ML for skill identification – Basic setup

- Language: Written & Spoken
- Cognitive Skills
- Functional Skills
- Personality
- Practical intelligence/Soft skills

Input programs

↓

ML magic!

↓

Prediction → Ground truth

rubric

Grading programs

Every TA's nightmare

Grading programs

A 2-3 hour fling for test takers with no *objective* feedback to improve

Grading programs

Helps companies find talent at scale

KHAN ACADEMY

coursera

UDACITY

MOOCs want a neat solution

Preliminaries

- We want to predict certain values y_i
- We have certain input values x_{ij}
- Simple example: Given a movie, predict how much business it will do
  - Inputs: actors, theme, director, time of year, etc.

Preliminaries

- In general, for n variables and m equations:

$$x_{11}a_1 + .. + x_{1n}a_n = y_1$$
$$x_{21}a_1 + .. + x_{2n}a_n = y_2$$
$$\dots$$
$$x_{m1}a_1 + .. + x_{mn}a_n = y_m$$

x_{ij} : Values we know

a_j : Constants we do not know

y_i : Values we want the constants to multiply and produce

- Two variables and two equations:

$$2p + 4q = 6$$
$$3p + 5q = 8$$

$$p = 1, q = 1$$

Preliminaries

- In general, for n variables and m equations:

$$x_{11}a_1 + .. + x_{1n}a_n = y_1$$
$$x_{21}a_1 + .. + x_{2n}a_n = y_2$$
$$\ldots$$
$$x_{m1}a_1 + .. + x_{mn}a_n = y_m$$

x_{ij} : Values we know

a_j : Constants we do not know

y_i : Values we want the constants to multiply and produce

<u>Remember Cramer's rule</u>

if $m > n$, then there exists no solution!

Input programs

↓

ML magic!

↓

Prediction → Ground truth

- In general, for n variables and m equations:

$$x_{11}a_1 + .. + x_{1n}a_n = y_1$$
$$x_{21}a_1 + .. + x_{2n}a_n = y_2$$
$$...$$
$$x_{m1}a_1 + .. + x_{mn}a_n = y_m$$

We don't fit exactly, we assume noise.

OLE formulation

Guess a_is such that the sum of square differences b/w the values of ys "predicted" and actual ys is minimum

$$\min_{\boldsymbol{a}} \sum_{1}^{m} (y_i - \boldsymbol{x}_i^T \boldsymbol{a})^2$$

Input programs

ML magic!

Prediction → Ground truth

- In general, for n variables and m equations:

$$x_{11}a_1 + .. + x_{1n}a_n = y_1$$
$$x_{21}a_1 + .. + x_{2n}a_n = y_2$$
$$\ldots$$
$$x_{m1}a_1 + .. + x_{mn}a_n = y_m$$

You have just learnt the most important and widely used machine learning technique!

Input

programs

↓

ML magic!

↓

Prediction → Ground truth

What does a grader look for?

OBJECTIVE

To print the pattern of integers

```
1
2  3
3  4  5
4  5  6  7
...
```

An implementation

```
void print_1(int N){
 for(i =1 ; i<=N; i++){
    print newline;
    count = i;
   for(j=0; j<i; j++)
     print count;
       count++;
  }
}
```

4. Is the inner loop dependent on the conditional of the first loop?

- a variable modified in the outer loop?

- a variable used in the conditional of the outer loop?

Score	Interpretation
5	**Completely correct and efficient** An efficient implementation of the problem using right control structures and data-dependencies.
4	**Correct with some inadvertent errors** Correct control structures and closely matching data-dependencies. Some silly mistakes fail the code to pass test-cases.
3	**Inconsistent logical structures** Right control structures start exist with few correct data dependencies
2	**Emerging basic structures** Appropriate keywords and tokens present, showing some understanding of the problem
1	**Gibberish code** Seemingly unrelated to problem at hand.

ML for skill identification – Basic setup

Input

programs

ML magic!

Prediction

Ground truth

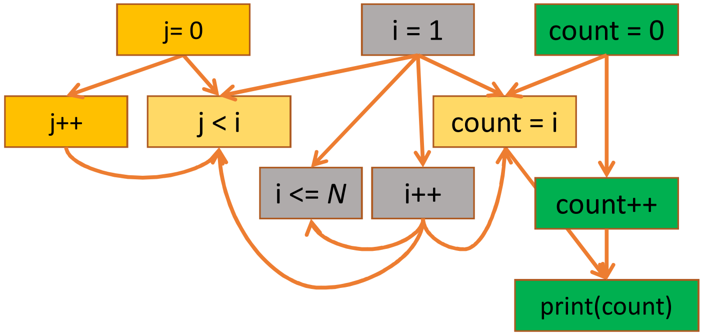
rubric

Features

TARGET PROGRAM

```
void print(int N){
for(i =1 ; i<=N; i++){
    print newline;
     count = i;
    for(j=0; j<i; j++)
      print count; count++;
    }
}
```

DATA DEPENDENCY GRAPH



Capture data-dependencies between expressions

– `i++` → `j < i` : **var (`i`) related to var (`j`) – previously incremented**

Key learning 4

When facing abstract inputs (like programs), think of features as counts

Aligning features

	Variables:**1**; operator: % ; DS: 0; const: **‘2’**	Variables:**2**; operator: < ; DS: 0; const: **0** #input	Variables:**1**; operator: ! = ; DS: 0; const: **‘a’**	Variables:**3**; operator: + ; DS: 1; const:0 @ loop_condition	`loop(loop (print)))`	
P-1	**1**	**2**	**0**	**0**	**0**	**good**
P-2	**0**	**4**	**3**	**0**	**0**	**bad**
...	**...**	**...**	**...**	**...**	**...**	**...**
P-M	**0**	**0**	**0**	**5**	**6**	**good**

Aligning features

Variables:**1**; operator: % ; DS: 0; const: **'2'**

Variables:**2**; operator: < ; DS: 0; const: **0** #input

Variables:**1**; operator: ! = ; DS: 0; const: **'a'**

Variables:**3**; operator: + ; DS: 1; const:0 @ loop_condition

`loop(loop(`**`print`**`))`

$$a_1 \begin{vmatrix} 1 \\ 0 \\ \dots \\ 0 \end{vmatrix} + a_2 \begin{vmatrix} 2 \\ 4 \\ \dots \\ 0 \end{vmatrix} + a_3 \begin{vmatrix} 0 \\ 3 \\ \dots \\ 0 \end{vmatrix} + a_4 \begin{vmatrix} 0 \\ 0 \\ \dots \\ 5 \end{vmatrix} + a_5 \begin{vmatrix} 0 \\ 0 \\ \dots \\ 6 \end{vmatrix} = \begin{matrix} \text{good} \\ \text{bad} \\ \dots \\ \text{good} \end{matrix}$$

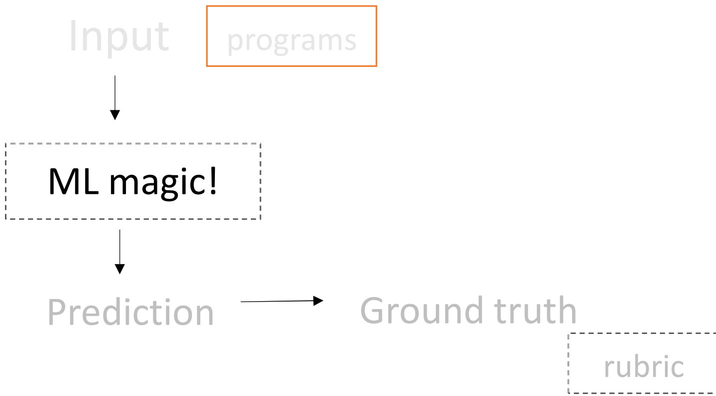
x_{ij}

ML for skill identification – Basic setup

- No magic – off the shelf regression libraries

- We get a correlation of ~0.9 with the expert labels!

- Inter-rater correlation is also about 0.9



Other problems

SVAR

- Automated fluency, pronunciation and other speech features
- Back-office industry

Motor skills grading

- How can fine motor skills be measured?
- Blue-collared skills

Personality instruments

- Based on the big-5 traits
- Effective in sales and other similar job roles

Skill-demand maps

- Interactive maps to see which skills are the hottest, sliced by geography
- Democraticizing skill-salary link

What does all the data say?

Variable	coefficient	p-value	Unit of change*	Odds (e^(coefficient*unit))
English score	**0.0026**	**0.00**	**100**	**1.29**
Quantitative Ability score	0.0003	0.38	100	1.03
Logical Ability score	**0.0014**	**0.01**	**100**	**1.15**
Domain Percentile	**0.0037**	**0.04**	**10**	**1.04**
10th class percentage	0.0083	0.16	10	1.09
12th class percentage	-0.0086	0.08	10	0.92
College Percentage	**0.0151**	**0.01**	**10**	**1.16**
Gender	-0.0442	0.60	1	0.96
Tier of college	**-0.1270**	**0.03**	**1**	**0.88**
Branch of study	**0.1515**	**0.05**	**1**	**1.16**
Tier of city	-0.0026	0.96	1	1.00
Openness to Experience score	-0.0253	0.58	1	0.98
Extraversion + Agreeableness Score	0.0001	1.00	1	1.00
Polychronicity score	0.0175	0.66	1	1.02
Constant	-4.1389	0.00		

Merit

A candidate with an AMCAT English & Logical score higher by 100 points each and domain percentile up by 10 points has 54% higher odds to get a job.

Bias

A candidate from a tier 2 campus has 12% (25%) lower odds and tier 3 campus has 24% (33%) lower odds to get a job even if he/she has the equal merit.

aspiringminds
Employability Quantified

Other fun stuff – research.aspiringminds.com

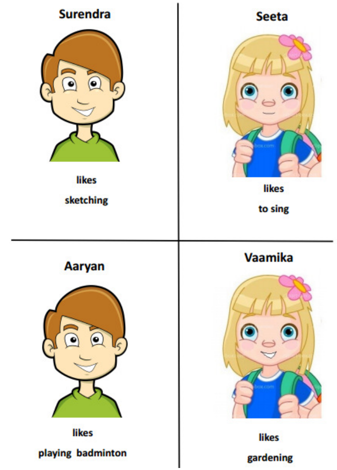
Data science for kids!

We taught 6th-9th graders DS!

4 cities covered; including UIUC!

Volunteer and organize your own camp

http://datasciencekids.org



ML India

ML-India

Track the latest in ML in India

Read prof. interviews

Attend such meetups

Volunteer and organize your own chapters

http://ml-india.org

aspiringminds
Employability Quantified