Towards More Adequate Value-Added Teacher Assessments: How Intervals Can Help

Karen Villaverde[1] and Olga Kosheleva[2]

[1]Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
kvillave@cs.nmsu.edu

[2]Department of Teacher Education
University of Texas at El Paso
El Paso, TX 79968, USA
olgak@utep.edu

1. Assessment is Important

- *Objective:* improve the efficiency of education.
- *Important:* to assess this efficiency, i.e., to describe this efficiency in quantitative terms.
- This is important on all education levels:
  - elementary schools
  - middle schools
  - high schools
  - universities
- Quantitative description is needed because
  - it allows natural comparison of different strategies of teaching and learning
  - and selection of the best strategy.

2. Need for Value-Added Assessment

- *Traditional assessment:* by the amount of knowledge that the students have after taking this class.
- *Example:* the average score of the students on some standardized test.
- *Comment:* this is actually how the quality of elementary/high school classes is now estimated in the US.
- *Limitation:* the class outcome depends
  - not only on the quality of the class, but
  - also on how prepared were the students when they started taking this class.
- *A more adequate assessment* should estimate the *added value* that the class brought to the students.

3. Current Approaches to Value-Added Assessment and their Limitations

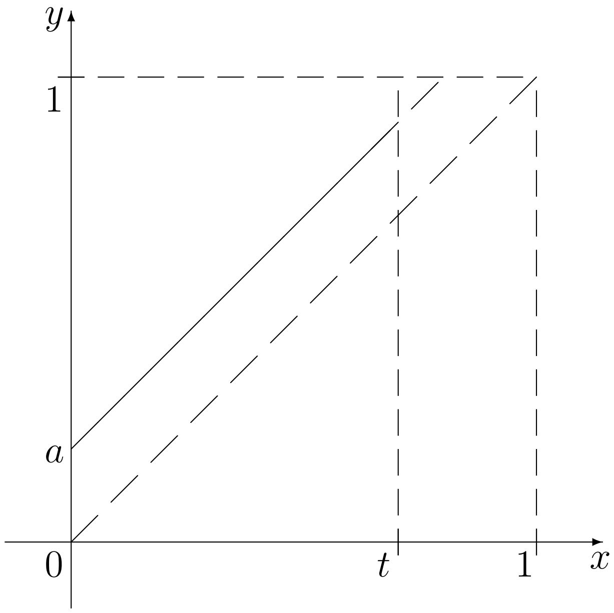
- *Main idea:* subtracting the outcome from the input.
- *Example:* subtract
  - the average grade after the class (on the post-test)
  - the average grade on similar questions asked before the class (on the pre-test).
- *Comment:* the existing techniques take into account additional parameters influencing learning.
- *Main limitation:* actually, the amount of knowledge learned depends on the initial knowledge.
- *Additional limitation:* the assessment values come from grading, and are therefore somewhat imprecise.

4. Natural Idea: Using Interval Techniques

- *Reminder:* assessments are imprecise, we usually only know bounds on the actual amount of knowledge.
- *Conclusion:* it is natural to use interval techniques to process the corresponding values.
- *In this paper:* we describe how to the use interval techniques.
- *Result:* interval techniques help us overcome both limitations of the existing value-added assessments.

5. Traditional Approach: Reminder

- *Reminder:* the post-test result y depends on the pre-test result x as $y \approx x + a$:

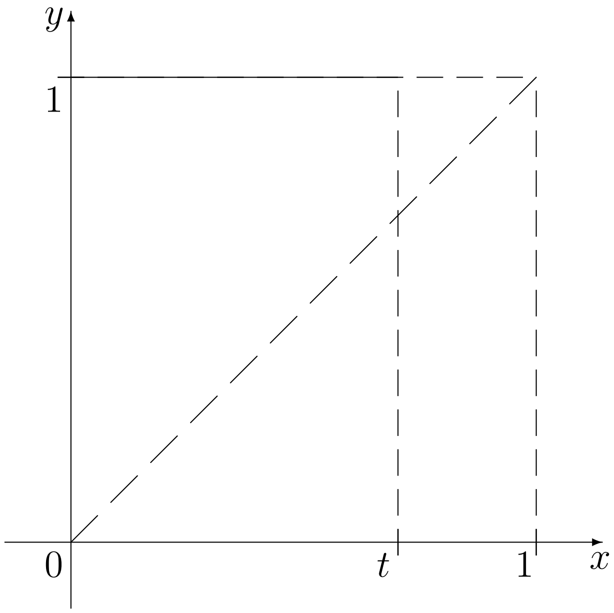


6. Linear Dependence instead of Addition: Idea

- *Problem:* the difference $y - x$ actually changes with x.
- *Natural next approximation:* $y \approx m \cdot x + a$.
- *Observation:* for f-s $f_1(x) = m_1 \cdot x + a_1$ and $f_2(x) = m_2 \cdot x + a_2$ corr. to two teaching strategies, we may have
  - $f_1(x_1) < f_2(x_1)$ for some x_1 and
  - $f_1(x_2) > f_2(x_2)$ for some $x_2 > x_1$.
- *Interpretation:*
  - for weaker students, with prior knowledge $x_1 < x_2$, the second strategy is better, while
  - for stronger students, with prior knowledge $x_2 > x_1$, the first strategy is better.
- *Conclusion:* the new model provides a more nuanced comparison between different teaching strategies.

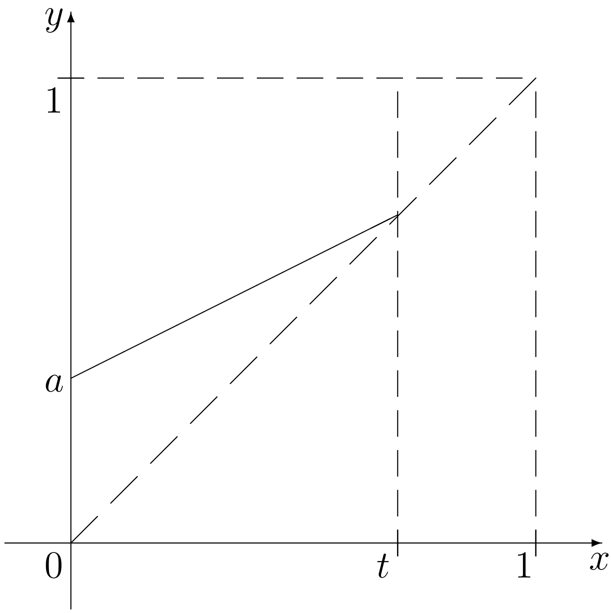
7. Ideal Case: Perfect Learning

- *Ideal case:* no matter what the original knowledge is, the resulting knowledge is perfect, $y \equiv 1$; then $m = 0$.



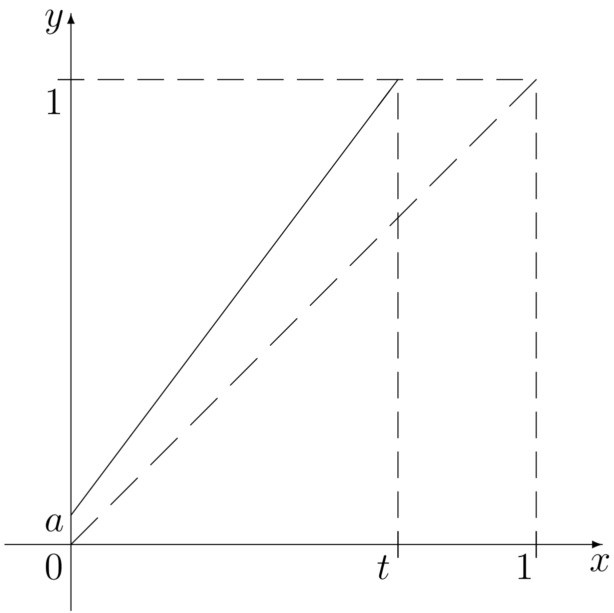
8. Example 2: Minimizing Failure Rate

- *Main idea:* to avoid failure, we concentrate on the students with low x; then $f(x) = m \cdot x + a$, with $m < 1$.



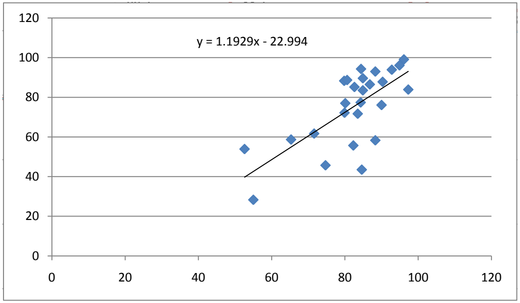
9. Example 3: Emphasis on Strong Students

- *Idea:* concentrate most of the effort on top students.
- *Result:* $f(x) = m \cdot x + a$, with $m > 1$.



10. How to Determine the Coefficients m and a: Ideal Case of Crisp Estimates

- *We know:* pre-test grades $x_1, \ldots, x_n$ and post-test grades $y_1, \ldots, y_n$.
- *Problem:* find m and a for which $y_i \approx m \cdot x_i + a$.
- *Least Squares method:* $\sum\limits_{i=1}^{n} (y_i - (m \cdot x_i + a))^2 \to \min\limits_{m,a}$.



11. Case of Interval Uncertainty: Analysis

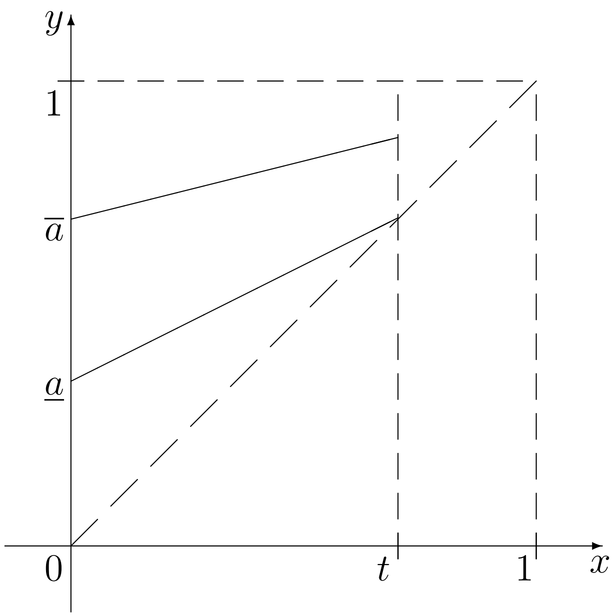
- *Fact:* the grade depends on assigning partial credit for partly correct solutions.
- *Known:* partial credit is somewhat subjective.
- *How to avoid this subjectivity:* letter grades such as A (corresponding to 90 to 100) are more objective.
- *Conclusion:* instead of the exact grade x_i, we have an interval $\mathbf{x} = [\underline{x}_i, \overline{x}_i]$ of possible grades.
- *Value-added assessment:* describe the dependence $\mathbf{y} = f(\mathbf{x})$ of the outcome grade $\mathbf{y}$ on the input grade $\mathbf{x}$:
  - we consider all the students for whom the input grade is within the interval $\mathbf{x}$;
  - then, $\mathbf{y} = f(\mathbf{x})$ is the set of all possible outcome grades for these students.

12. Which Interval-to-Interval Functions Are Reasonable

- *Example:* suppose that
  - when the pre-test grade x is in $\mathbf{x}_1 = [80, 90]$, then the post-test grade y is in $\mathbf{y}_1 = f(\mathbf{x}_1) = [85, 95]$;
  - when $x \in \mathbf{x}_2 = [90, 100]$, then $y \in \mathbf{y}_2 = f(\mathbf{x}_2) = [92, 100]$.
- *Argument:* when $x \in \mathbf{x}_1 \cup \mathbf{x}_2$, then $x \in \mathbf{x}_1$ or $x \in \mathbf{x}_2$, so $y \in \mathbf{y}_1$ or $y \in \mathbf{y}_2$.
- *Conclusion:* $f(\mathbf{x}_1 \cup \mathbf{x}_2) = f(\mathbf{x}_1) \cup f(\mathbf{x}_2)$.
- *Similar conclusion:* $f(\mathbf{x}) = \bigcup\limits_{x \in \mathbf{x}} f([x, x])$.
- *Notation:* $[\underline{f}(x), \overline{f}(x)] \stackrel{\text{def}}{=} f([x, x])$.
- *Result:* all reasonable functions $f(\mathbf{x})$ have the form $f([\underline{x}, \overline{x}]) = [\underline{y}, \overline{y}]$, where $\underline{y} \stackrel{\text{def}}{=} \min\limits_{x \in [\underline{x}, \overline{x}]} \underline{f}(x)$; $\overline{y} \stackrel{\text{def}}{=} \max\limits_{x \in [\underline{x}, \overline{x}]} \overline{f}(x)$.

13. Case of Interval Uncertainty: Algorithm

- *Idea:* based on $[\underline{x}_i, \overline{x}_i]$ and $[\underline{y}_i, \overline{y}_i]$, we use Least Squares to find values s.t. $\underline{y}_i \approx \underline{m} \cdot \underline{x}_i + \underline{a}$ and $\overline{y}_i \approx \overline{m} \cdot \overline{x}_i + \overline{a}$.



A. Appendix: Case of Fuzzy Uncertainty

- *Interval assumption:* we assumed that the interval $[\underline{x}, \overline{x}]$ is guaranteed to contain the actual (unknown) value x.
- *In reality:* the bounds that we know are "fuzzy", i.e., they contain x only with some degree of confidence α.
- *Conclusion:* we have different intervals $[\underline{x}(\alpha), \overline{x}(\alpha)]$ corresponding to different degrees α.
- *Observation:* this is equivalent to knowing a fuzzy set with given α-cuts $[\underline{x}(\alpha), \overline{x}(\alpha)]$.
- *Resulting algorithm:* for each α, we find the interval-values linear function

$$[\underline{m}(\alpha) \cdot x + \underline{a}(\alpha), \overline{m}(\alpha) \cdot x + \overline{a}(\alpha)]$$

corresponding to this α.

B. How to Use the Resulting Fuzzy Estimates to Compare Different Teaching Strategies

- From the input fuzzy grades $X_1, \ldots, X_n$, we extract α-cuts corresponding to their α-cuts $X_i(\alpha)$.
- We know input-output functions corresponding $f_j([\underline{x}, \overline{x}])$ corresponding to different strategies j.
- We apply these functions to intervals $X_i(\alpha)$ and get fuzzy estimates $Y_{1,j}, \ldots, Y_{n,j}$ for post-test results.
- For each j, we apply the objective function to values $Y_{1,j}, \ldots, Y_{n,j}$.
- Thus, we get the fuzzy estimate V_j of the quality of the j-th strategy.
- We then use fuzzy optimization techniques to select the teaching strategy with the largest value V_j.