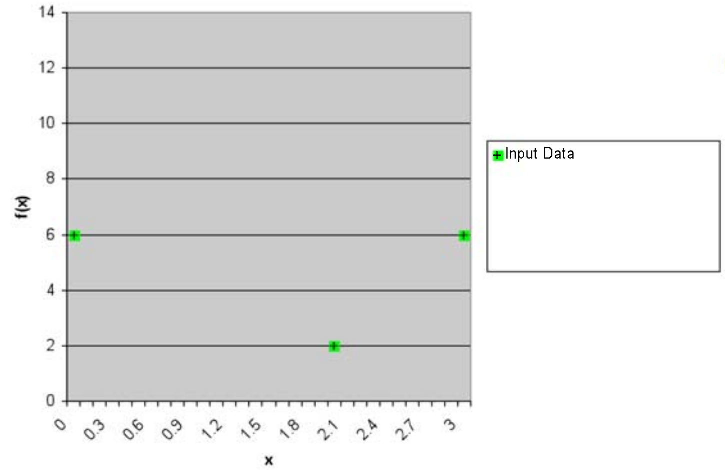
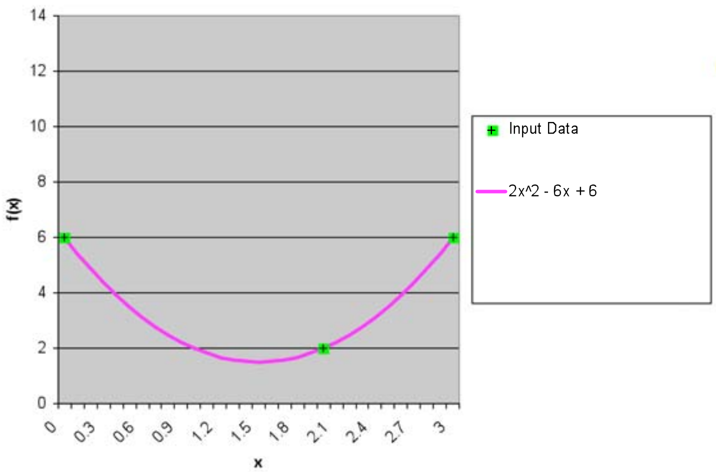
Learning in the Limit

Gold's Theorem

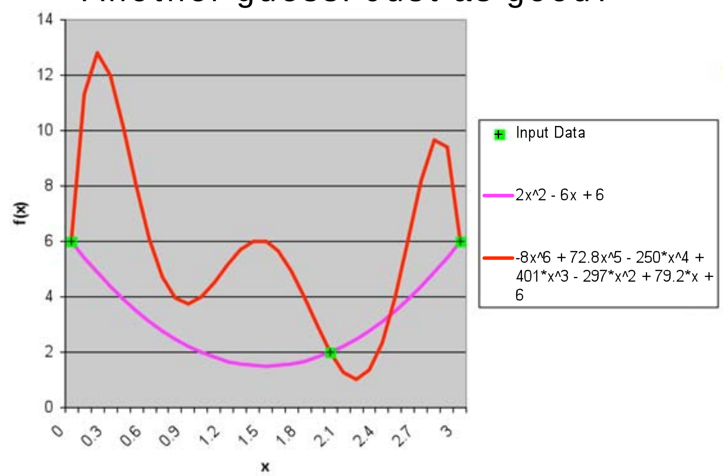
Observe some values of a function



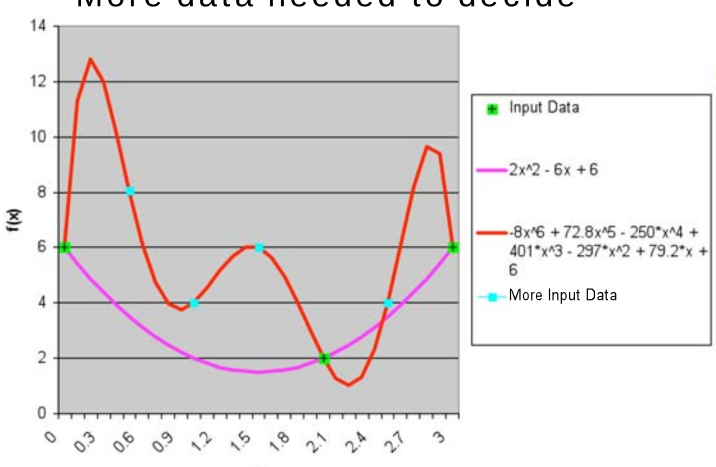
Guess the whole function



Another guess: Just as good?



More data needed to decide



Poverty of the Stimulus

- Never enough input data to completely determine the polynomial …
  - Always have infinitely many possibilities
- … unless you know the order of the polynomial ahead of time.
  - 2 points determine a line
  - 3 points determine a quadratic
  - etc.
- In language learning, is it enough to know that the target language is generated by a CFG?
  - without knowing the size of the CFG?

Language learning: What kind of evidence?

- Children listen to language [unsupervised]
- Children are corrected?? [supervised]
- Children observe language in context
- Children observe frequencies of language

Remember: Language = set of strings

Poverty of the Stimulus (1957)

Chomsky: Just like polynomials: never enough data unless you know something in advance. So kids must be born knowing what to expect in language.

- Children listen to language
- Children are corrected??
- Children observe language in context
- Children observe frequencies of language

Gold's Theorem (1967)

a simple negative result along these lines:
kids (or computers) can't learn much without supervision, inborn knowledge, or statistics

- Children listen to language
- Children are corrected??
- Children observe language in context
- Children observe frequencies of language

The Idealized Situation

- Mom talks
- Baby listens

- 1. Mom outputs a sentence
- 2. Baby hypothesizes what the language is (given all sentences so far)
- 3. Goto step 1

- Guarantee: Mom's language *is* in the set of hypotheses that Baby is choosing among
- Guarantee: Any sentence of Mom's language is *eventually* uttered by Mom (even if infinitely many)
- Assumption: Vocabulary (or alphabet) is finite.

Can Baby learn under these conditions?

- Learning in the limit:
  - There is some point at which Baby's hypothesis is correct and never changes again. Baby has converged!
  - Baby doesn't have to **know** that it's reached this point – it can keep an open mind about new evidence – but if its hypothesis is right, no such new evidence will ever come along.

- A class C of languages is **learnable in the limit** if one could construct a perfect C-Baby that can learn any language $L \in C$ in the limit from a Mom who speaks L.

- Baby knows the class C of possibilities, but not L.
- Is there a perfect finite-state Baby?
- Is there a perfect context-free Baby?

Languages vs. Grammars

- Does Baby have to get the right grammar?
- (E.g., does VP have to be called VP?)

- Assumption: Finite vocabulary.

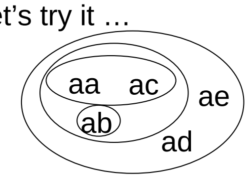
Conservative Strategy

- Baby's hypothesis should always be smallest language consistent with the data
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}

Mom	aa	ab	ac	ab	aa	…
Baby	L3	L1	L1	L1	L1	

Conservative Strategy

- Baby's hypothesis should always be smallest language consistent with the data
- Works for finite languages? Let's try it …
  - Language 1: {aa,ab,ac}
  - Language 2: {aa,ab,ac,ad,ae}
  - Language 3: {aa,ac}
  - Language 4: {ab}



Mom	aa	ab	ac	ab	aa	…
Baby	L3	L1	L1	L1	L1	

Evil Mom

- To find out whether Baby is perfect, we have to see whether it gets 100% even in the most adversarial conditions
- Assume Mom is trying to fool Baby
  - although she must speak only sentences from L
  - and she must eventually speak each such sentence
- Does Baby's strategy work?

An Unlearnable Class

- Class of languages:
  - Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
    - If the true language is L_∞, can Mom really follow rules?
    - Must eventually speak every sentence of L_∞. Possible?
    - Yes: ε; a, b; aa, ab, ba, bb; aaa, aab, aba, abb, baa, …
  - Our class is C = {L_0, L_1, … L_∞}

An Unlearnable Class

- Let L_n = set of all strings of length < n
  - What is L_0?
  - What is L_1?
  - What is L_∞?
- Our class is C = {L_0, L_1, … L_∞}
- A perfect C-baby will distinguish among all of these depending on the input.
- But there is no perfect C-baby …

An Unlearnable Class

- Our class is C = {L_0, L_1, … L_∞}
- Suppose Baby adopts conservative strategy, always picking smallest possible language in C.
- So if Mom's longest sentence so far has 75 words, baby's hypothesis is L_{76}.
- This won't always work: What language can't a conservative Baby learn?

An Unlearnable Class

- Our class is $C = \{L_0, L_1, \ldots L_\infty\}$
- Could a non-conservative baby be a perfect C-Baby, and eventually converge to any of these?
- **Claim:** *Any* perfect C-Baby must be "quasi-conservative":
  - If true language is L_{76}, and baby posits something else, baby must still eventually come back and guess L_{76} (since it's perfect).
  - So if longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually baby must actually return to the conservative guess L_{76}.
  - Agreed?

Mom's Revenge

If longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually a perfect C-baby must actually return to the conservative guess L_{76}.

- Suppose true language is L_∞.
- Evil Mom can prevent our supposedly perfect C-Baby from converging to it.
- If Baby ever guesses L_∞, say when the longest sentence is 75 words:
  - Then Evil Mom keeps talking from L_{76} until Baby capitulates and revises her guess to L_{76} – as any perfect C-Baby must.
  - So Baby has *not* stayed at L_∞ as required.
- Then Mom can go ahead with longer sentences. If Baby ever guesses L_∞ again, she plays the same trick again.

Mom's Revenge

If longest sentence so far is 75 words, and Mom keeps talking from L_{76}, then eventually a perfect C-baby must actually return to the conservative guess L_{76}.

- Suppose true language is L_∞.
- Evil Mom can prevent our supposedly perfect C-Baby from converging to it.
- If Baby ever guesses L_∞, say when the longest sentence is 75 words:
  - Then Evil Mom keeps talking from L_{76} until Baby capitulates and revises her guess to L_{76} – as any perfect C-Baby must.
  - So Baby has *not* stayed at L_∞ as required.
- Conclusion: There's no perfect Baby that is guaranteed to converge to L_0, L_1, … or L_∞ as appropriate. If it always succeeds on finite languages, Evil Mom can trick it on infinite language.

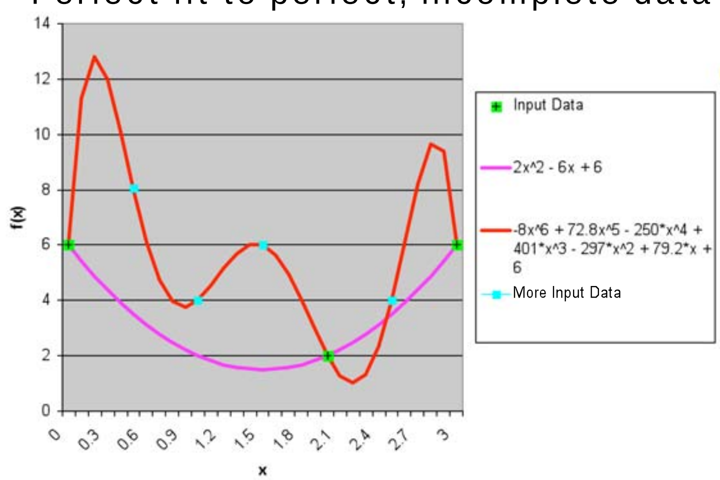
Implications

- We found that $C = \{L_0, L_1, \ldots L_\infty\}$ isn't learnable in the limit.
- How about class of finite-state languages?
  - Not unless you limit it further (e.g., # of states)
  - After all, it includes all languages in C, and more, so learner has harder choice
- How about class of context-free languages?
  - Not unless you limit it further (e.g., # of rules)

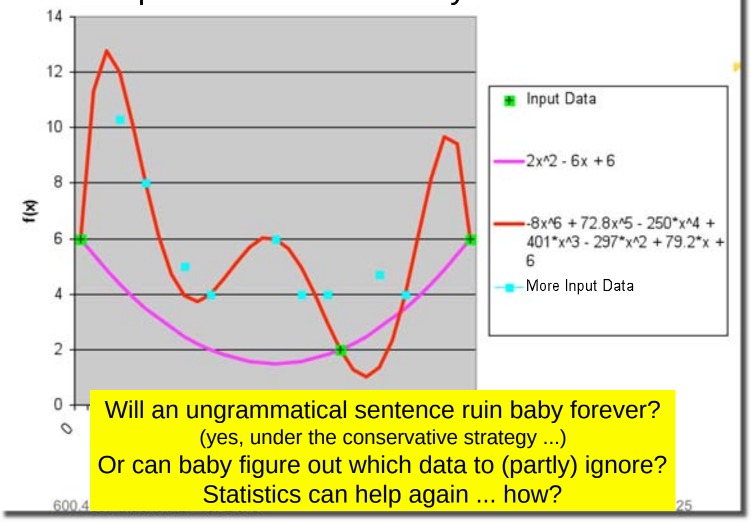
Punchline

- But class of *probabilistic* context-free languages *is* learnable in the limit!!
- If Mom has to output sentences randomly with the appropriate probabilities,
  - she's unable to be too evil
  - there are then perfect Babies that are guaranteed to converge to an appropriate probabilistic CFG
- I.e., from hearing a finite number of sentences, Baby can correctly converge on a grammar that predicts an infinite number of sentences.
  - Baby is generalizing! Just like real babies!

Perfect fit to perfect, incomplete data



Imperfect fit to noisy data



Will an ungrammatical sentence ruin baby forever?

(yes, under the conservative strategy ...)

Or can baby figure out which data to (partly) ignore?

Statistics can help again ... how?