Machine learning meta theorems

Oliver W. Layton

CS251: Data analysis and visualization

Lecture 30, Spring 2019

Monday April 22

Plan

- No free lunch theorem
- Ugly duckling theorem

Which machine learning model is best?

- Is there an overall performance advantage to one machine learning model in all classification tasks (medical diagnosis, people identification, detecting spam email, etc.)?
- Can we always find a classifier that performs better than chance at ALL tasks?

Example: Learning binary features

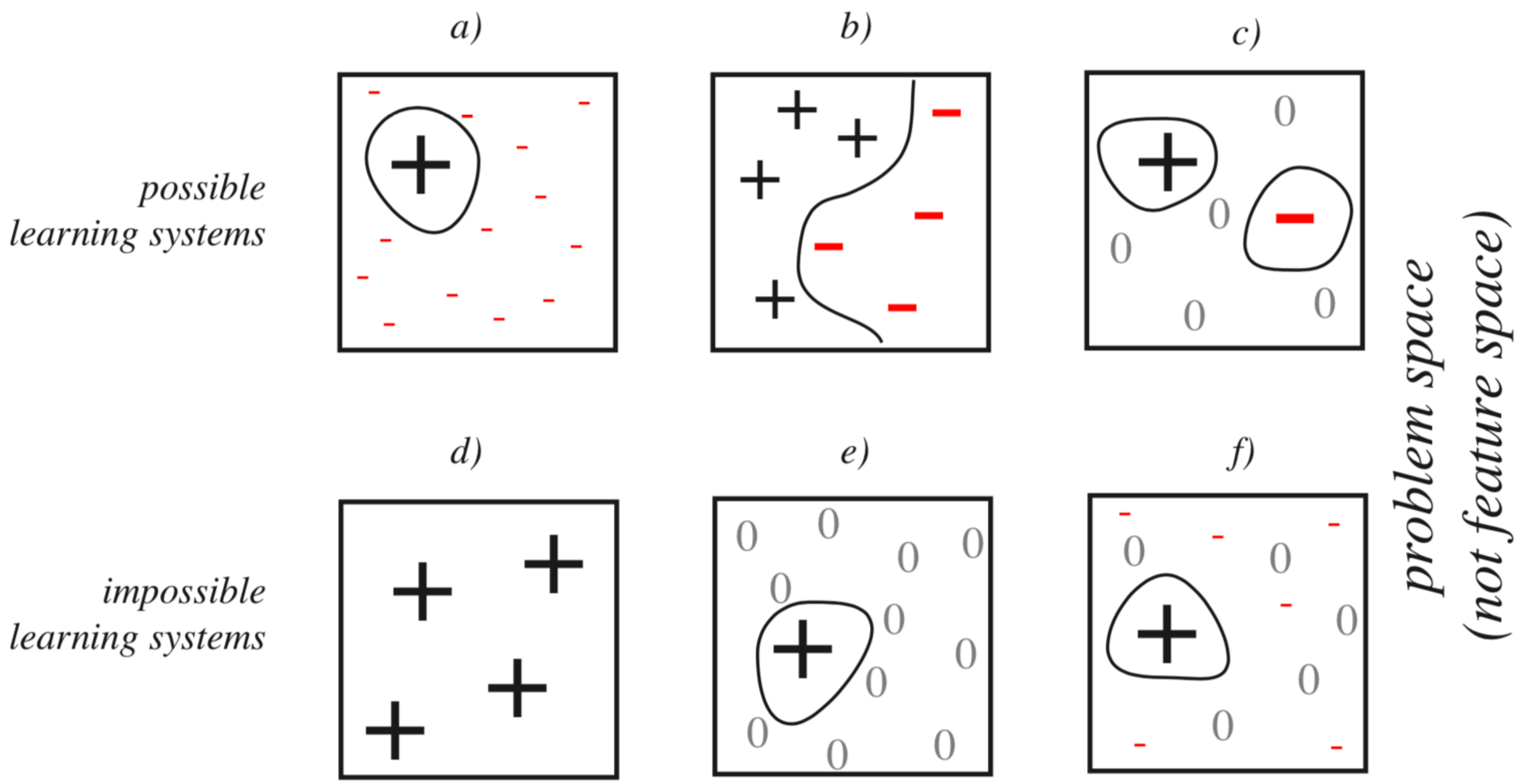
	features	class	Model1 Prediction	Model2 Prediction
training	000	+1	+1	+1
	001	-1	-1	-1
	010	+1	+1	+1
test	111	-1	+1	-1
	011	+1	+1	-1
	100	-1	+1	-1
	101	+1	+1	-1
	110	+1	+1	-1

No Free Lunch Theorem

- According to the **No Free Lunch Theorem**.
- Sadly, there is **no** classifier that performs better than chance at ALL tasks or is inherently superior.
- There is no context- or problem-independent reason to favor one machine learning method over another.
- The apparent superiority may be due to the selection of your data from a larger potential "pool".

No Free Lunch Theorem summary

If a machine learning model performs "above average" over some set of problems, then it must perform "worse than average" elsewhere.



'+' means above average performance, 0 means average, - means worse than average.

No Free Lunch Theorem Upshot

- For specific problem domains, consider range of machine learning algorithms.
- Take advantage of prior information, domain knowledge, data distribution, and amount of data.

Ugly Duckling Theorem

- Without prior knowledge, is there a principled way to determine whether two distinct patterns are more similar ("closer") than two other distinct patterns?
- Example: Binary features for blind in right eye (10) and blind in left eye (01).
  - Blind in right (left) eye **more similar** to totally blind person and normally sighted person (1 bit flip) **than** to left (right) blind.
  - Not best for all situations (e.g driving capabilities).

Example: Different feature representations

	f1	**f2**	**f1'**	**f2'**
x1	0	0	0	1
x2	0	1	0	0
x3	1	0	1	0
x4	1	1	1	1

- f1 (Blind in left eye), f2 (Blind in right eye) dimensions.
- f1' (Blind in left eye), f2' (same in both eyes) dimensions.
- No general advantage of one coding scheme vs. another.

Example: Ugly duckling Summary

- Notion of similarity between patterns depends implicitly on assumptions which may or may not be correct.
- In the absence of assumptions, no general advantage of one coding scheme vs. another.
- What do you think?
- There is no silver bullet method that will work well in all data sets.
- You must choose your algorithms carefully and pay attention to the details (i.e. choose an appropriate distance metric).