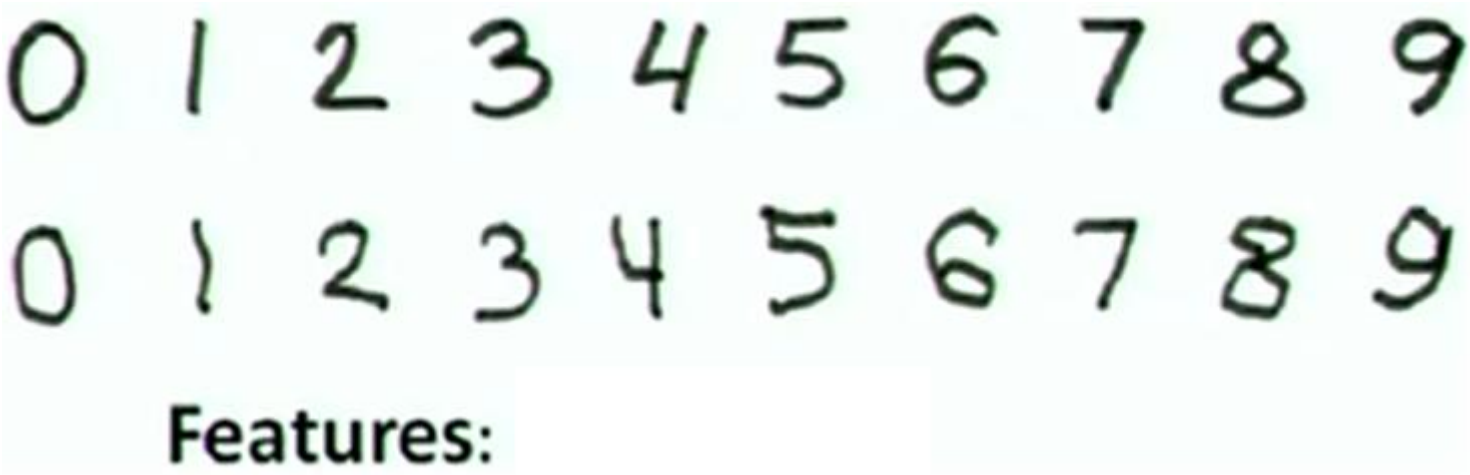
Multiclass

Nazerfard, Ehsan
nazerfard@aut.ac.ir

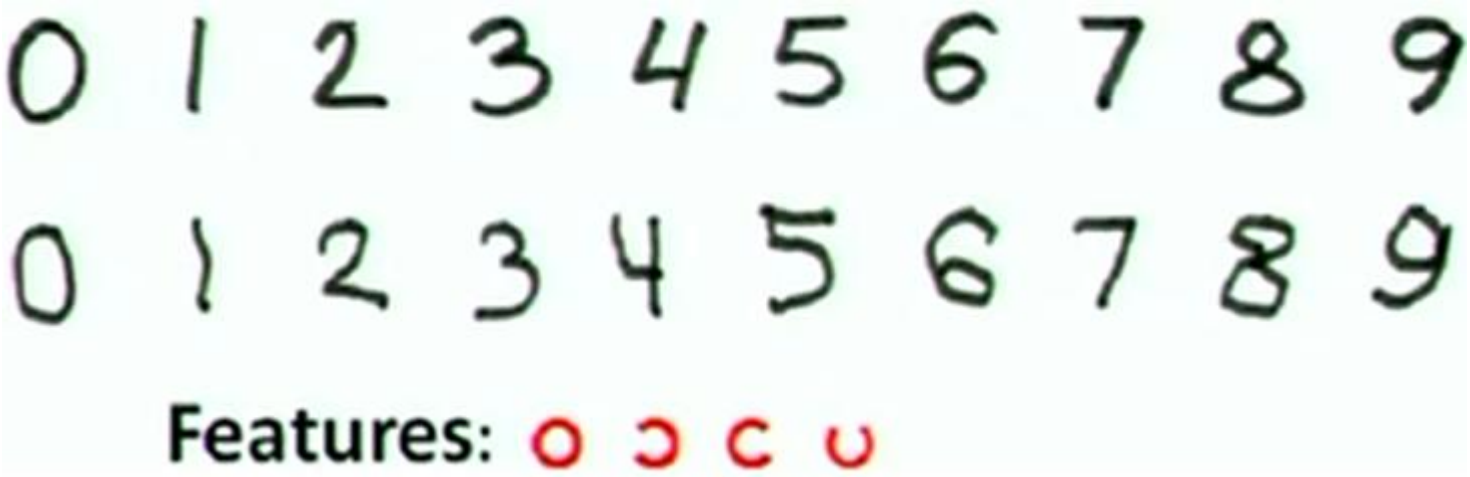
Topics

- Multiclass prediction
  - Decision Tree example
- One vs. All
- Non-Parametric Models
- No free lunch theorem
- UCI machine learning repository

Handwriting Digit Recognition



Multiclass Prediction (example)

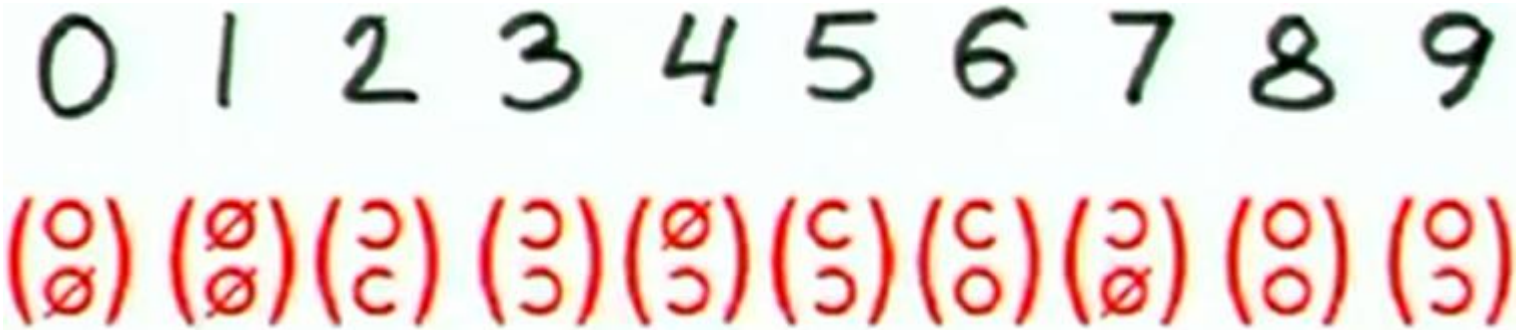


Example © Ron Bekkerman, Scaling up Machine Learning, 2011

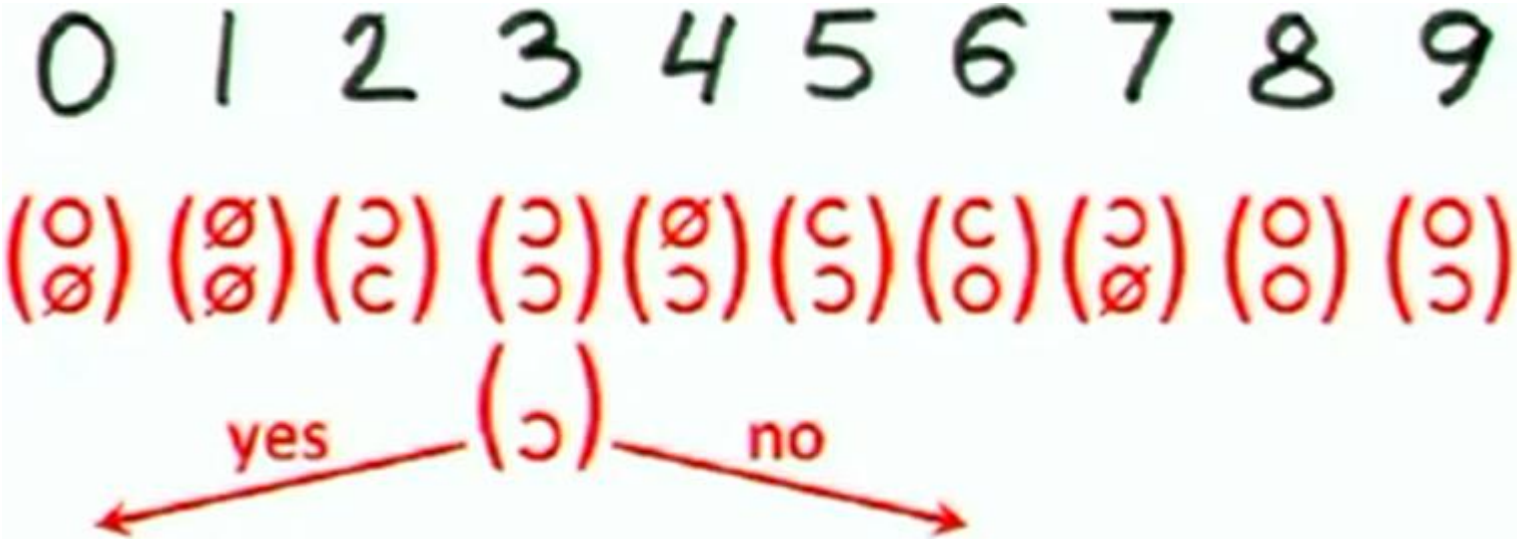
Feature Set

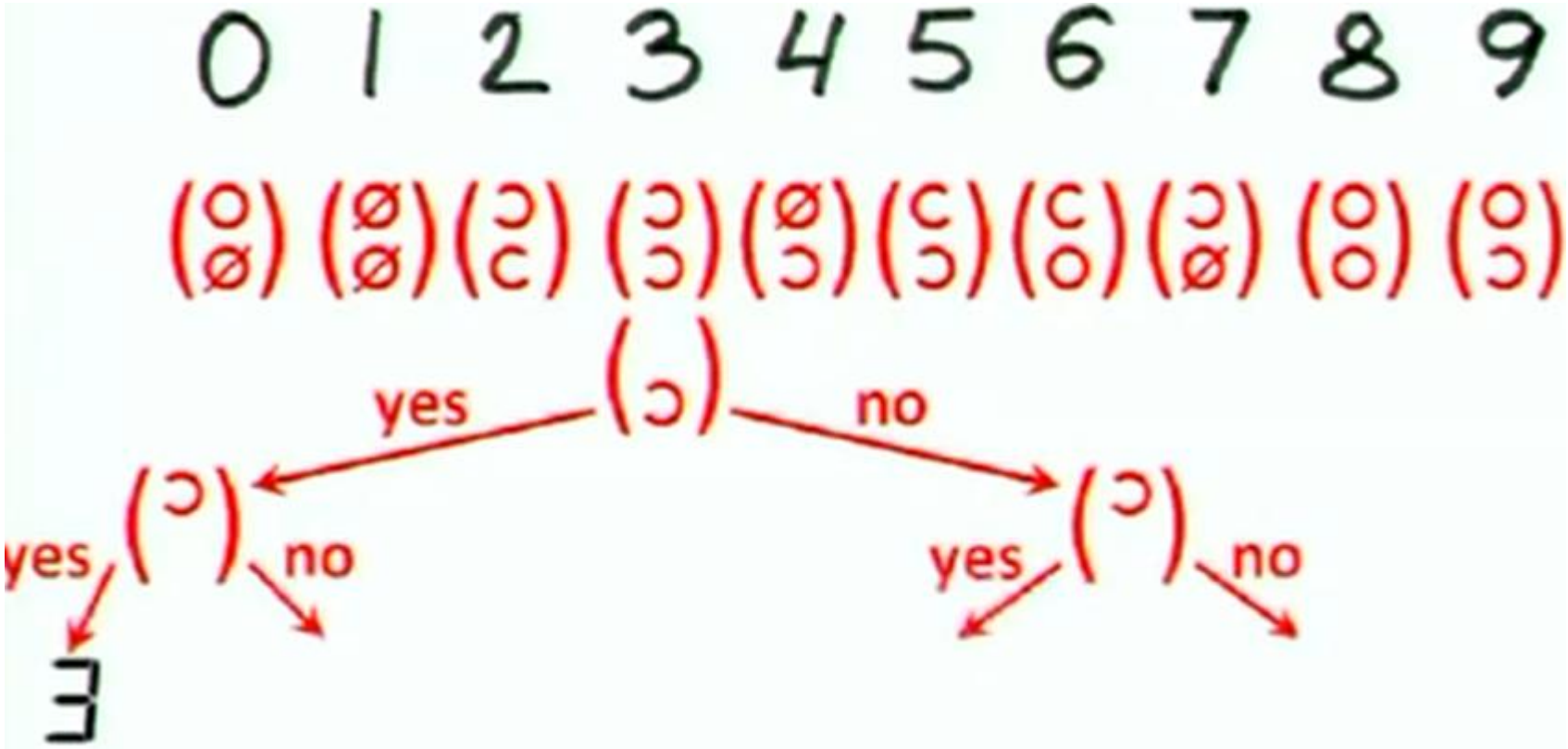
Example © Ron Bekkerman, Scaling up Machine Learning, 2011

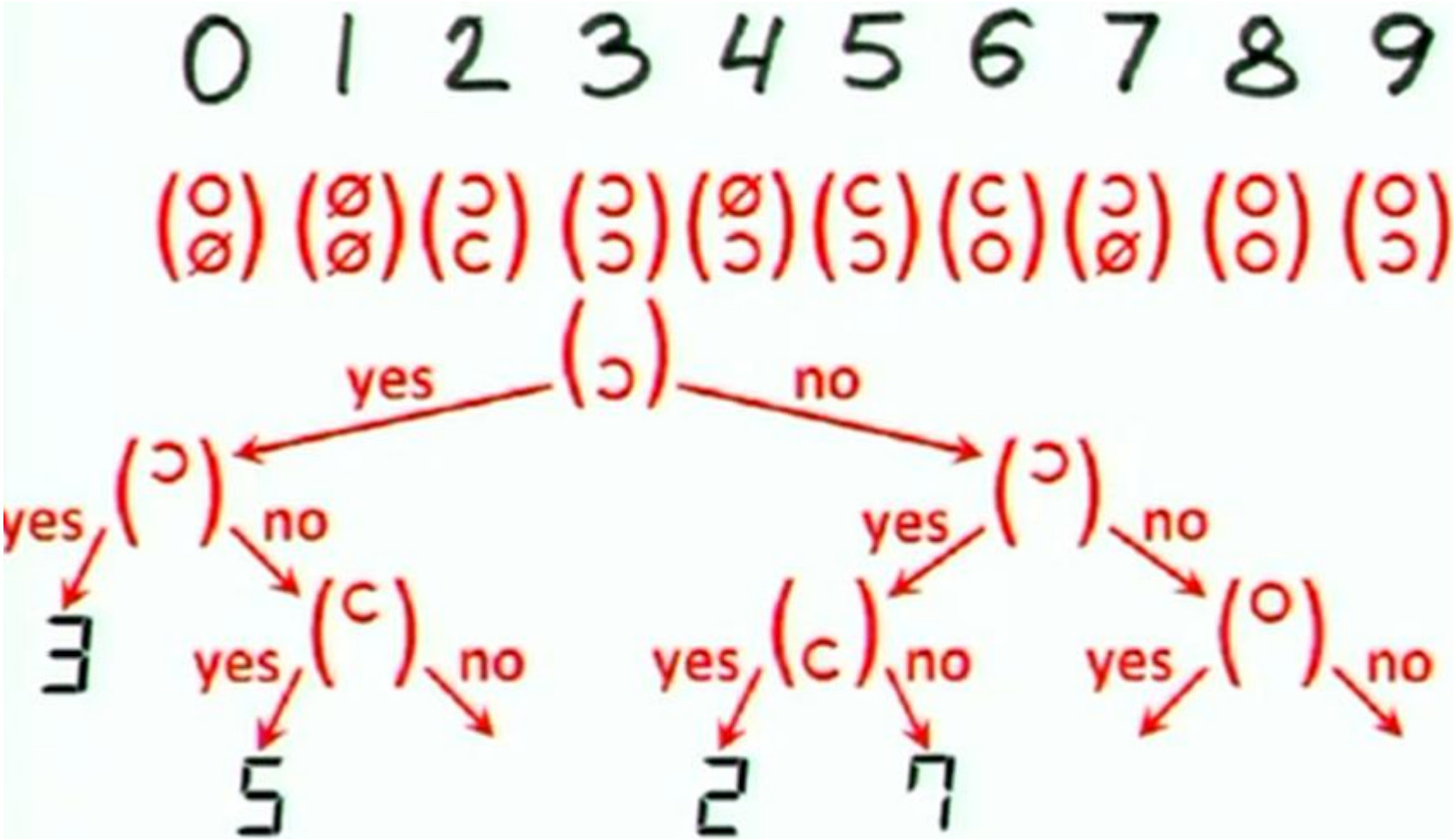
Feature Set (cont.)

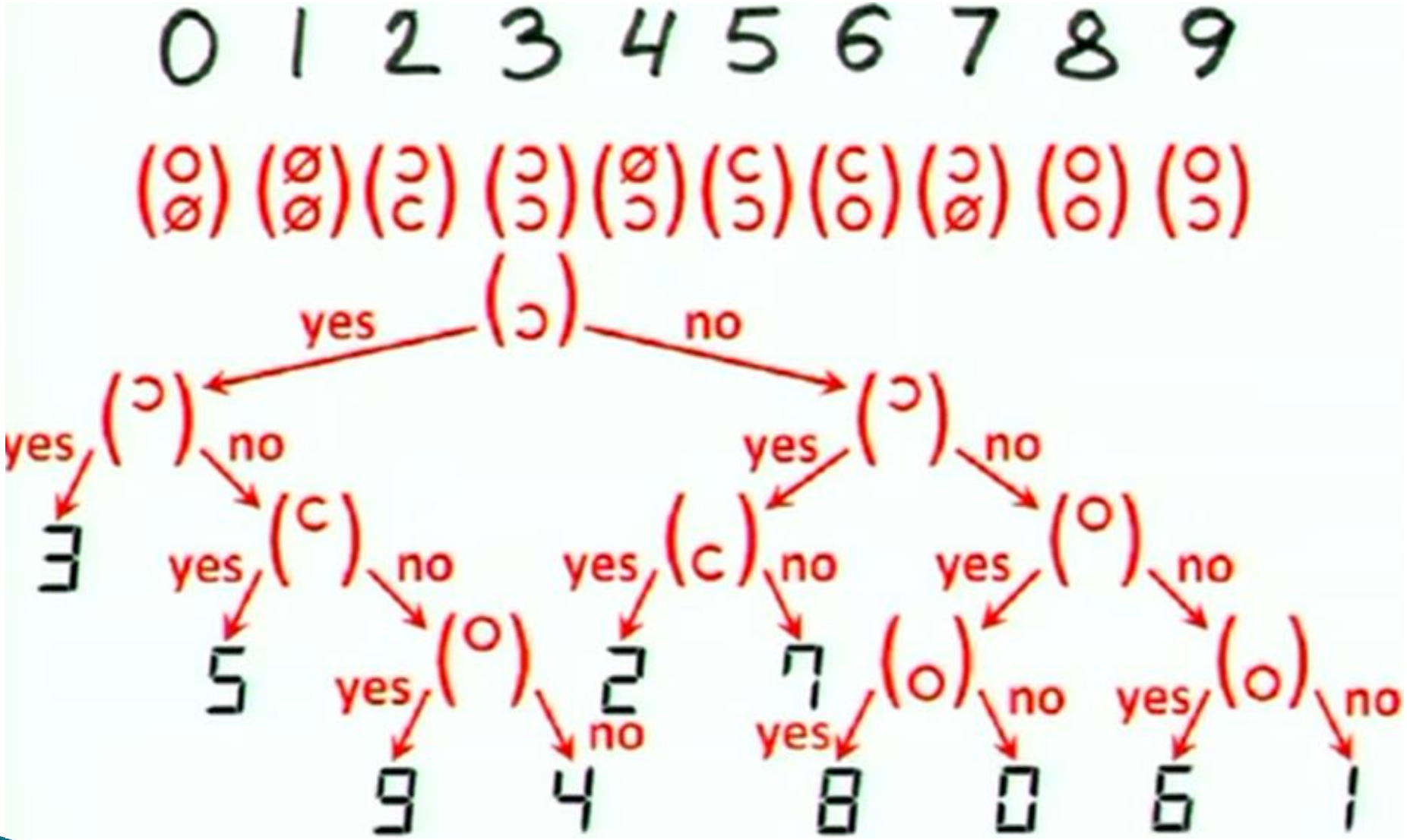


Decision Tree









Confusion Matrix

- A multiclass (A, B, …, E) prediction:

		Predicted				
		A	B	C	D	E
Actual	A					
	B					
	C					
	D					
	E					

Confusion Matrix (cont.)

- A multiclass prediction:

		Predicted				
		A	B	C	D	E
Actual	A	$\#TP_A$	$\#E_{AB}$	$\#E_{AC}$	$\#E_{AD}$	$\#E_{AE}$
	B	$\#E_{BA}$	$\#TP_B$	.	.	.
	C	$\#E_{CA}$	.	$\#TP_C$	.	.
	D	$\#E_{DA}$	.	.	$\#TP_D$	.
	E	$\#E_{EA}$	.	.	.	$\#TP_E$

Confusion Matrix (cont.)

- A multiclass prediction:

		Predicted					
		A	B	C	D	E	Accuracy
Actual	A	TP_A	E_{AB}	E_{AC}	E_{AD}	E_{AE}	–
	B	E_{BA}	TP_B	.	.	.	–
	C	E_{CA}	.	TP_C	.	.	–
	D	E_{DA}	.	.	TP_D	.	–
	E	E_{EA}	.	.	.	TP_E	–

Performance Metrics

- How to Compute the followings (one vs. all):
  - Accuracy
  - True Positive (TP)
  - False Negative (FN)
  - False Positive (FP)
  - True Negative (TN)
  - Sensitivity
  - Precision
  - …

Non-Parametric Models

- A non-parametric model
  - Does it refer to a model without parameters?
  - Examples:
    - K-nearest neighbors
    - Decision Tree
    - …

Miscellaneous

- No free lunch (NFL) theorem
  - The theorem states that no learner can succeed on all learnable tasks – every learner has tasks on which it fails while other learners succeed [1].

- UCI machine learning repository
  - https://archive.ics.uci.edu/ml/

Further Reading

1. Understanding Machine Learning: From Theory to Algorithms, Shai Shalev-Shwartz and Shai Ben-David, Cambridge University Press, 2014.

   (Available online – No for distribution)