1) BAYES OPTIMAL CLASSIFIER

2) INDUCTIVE BIAS

3) LIMITS OF LEARNABLE

4) UNDERFITTING VS. OVERFITTING

5) TRAINING VS. TEST DATA

6) PARAMETERS VS. HYPER-PARAMETERS

7) ML IN THE REAL WORLD

8) FURTHER READING

1) BAYES OPTIMAL CLASSIFIER

Given distribution $D \sim (x, y)$

$$f_{BO}(\bar{x}) = \arg\max_{y} D(x, y \mid x = \bar{x})$$

WHAT ABOUT PROBABILISTIC CASE?

$\hat{y}$ \ y	YES	NO	
YES	0.81	0.09	0.9
NO	0.09	0.01	0.1
	0.9	0.1	

2) INDUCTIVE BIAS

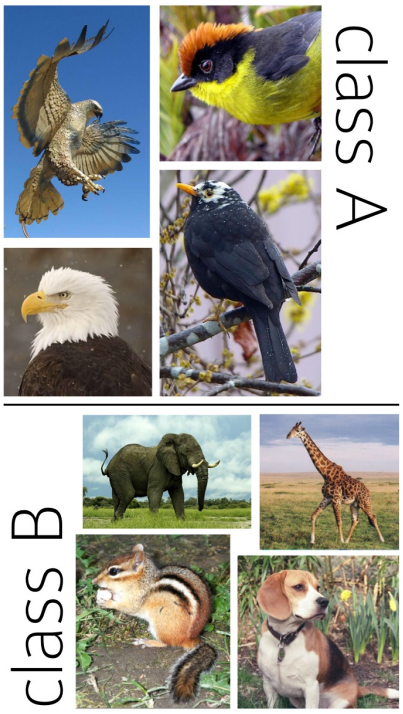

Figure 2.1: Training data for a binary classification problem.

3) LIMITS OF LEARNING

WHAT SOURCES OF NOISE?

WHAT ASSUMPTIONS ABOUT TASK?

- FEATURE SHIFT
- LABEL SHIFT
- DISTRIBUTION SHIFT
- DISTRIBUTION?

4) UNDER VS. OVERFITTING

5) TRAINING VS. TEST DATA

6) HYPER PARAMETERS

7) ML IN THE REAL WORLD

1	real world goal	increase revenue
2	real world mechanism	better ad display
3	learning problem	classify click-through
4	data collection	interaction w/ current system
5	collected data	query, ad, click
6	data representation	bow[2], $\pm$ click
7	select model family	decision trees, depth 20
8	select training data	subset from april'16
9	train model & hyperparams	final decision tree
10	predict on test data	subset from may'16
11	evaluate error	zero/one loss for $\pm$ click
12	deploy!	(hope we achieve our goal)

Figure 2.4: A typical design process for a machine learning application.

HOW DO WE DEBUG ANY OF THIS?

8) FURTHER READINGS

- "THE REUSABLE HOLDOUT"

- "ML : THE HIGH INTEREST CREDIT CARD OF TECHNICAL DEBT"