Learning from Data
Homework # 1

Khoi Pham

The Learning Problem

1. What types of learning, if any, best describe the following three scenarios:

(i) A coin classification system is created for a vending machine. In order to do this, the developers obtain exact coin specifications from the U.S. Mint and derive a statistical model of the size, weight, and denomination, which the vending machine then uses to classify its coins.
Analysis: this scenario describes a probabilistic design approach to classify different types of coins, thus it is not a learning approach.

(ii) Instead of calling the U.S. Mint to obtain coin information, an algorithm is presented with a large set of labeled coins. The algorithm uses this data to infer decision boundaries which the vending machine then uses to classify its coins.
Analysis: in this scenario, it is specified that all the obtained coins are already labeled and the system shall infer the decision boundaries from these labeled data. Thus, it is supervised learning.

(iii) A computer develops a strategy for playing Tic-Tac-Toe by playing repeatedly and adjusting its strategy by penalizing moves that eventually lead to losing.
Analysis: this scenario exactly describes the process of how reinforcement learning works. Instead of labeling each move as good or bad, the computer is given a value describing how bad that move is.

Answer: [d]

2. Which of the following problems are best suited for the learning approach?

(i) Classifying numbers into primes and non-primes.
Analysis: this is not suited for learning, since we can pin down mathematically what a prime number is, and our current algorithm for checking prime number is already efficient.

(ii) Detecting potential fraud in credit card charges.
Analysis: this suited for learning.

(iii) Determining the time it would take a falling object to hit the ground.
Analysis: this is not suited, as we can calculate the desired time using equations in physics.

(iv) Determining the optimal cycle for traffic lights in a busy intersection. *Analysis:* this is suited for learning.

Answer: [a]

Bins and Marbles

3. We have 2 opaque bags, each containing 2 balls. One bag has 2 black balls and the other has a black ball and a white ball. You pick a bag at random and then pick one of the balls in that bag at random. When you look at the ball, it is black. You now pick the second ball from that same bag. What is the probability that this ball is also black?

We define the following events:

A: pick bag with 2 black balls.

B: pick bag with 1 black ball and 1 white ball.

C1: pick a black ball in the 1st turn.

C2: pick a black ball in the 2nd turn.

The problem asks to calculate the conditional probability of picking a black ball in the 2nd turn, based on the condition that in the 1st turn, we also picked a black ball, that is $P(C2 \mid C1)$.

$P(A) = P(B) = \frac{1}{2}$

$P(C1 \mid A) = 1$

$P(C1 \mid B) = \frac{1}{2}$

$P(C2 \mid C1, A) = 1$

$P(C2 \mid C1, B) = 0$

We have:

$$
\begin{aligned}
P(C2 \mid C1) &= P(C2 \cap A \mid C1) + P(C2 \cap B \mid C1) \\
&= P(C2 \mid A, C1)P(A \mid C1) + P(C2 \mid B, C1)P(B \mid C1) \\
&= P(A \mid C1) \\
&= \frac{P(C1 \mid A)P(A)}{P(C1)} \\
&= \frac{1/2}{P(C1 \mid A)P(A) + P(C1 \mid B)P(B)} \\
&= \frac{1/2}{1/2 + 1/2 * 1/2} \\
&= \frac{2}{3}
\end{aligned}
$$

Answer: [d]

Consider a sample of 10 marbles drawn from a bin that has red and green marbles. The probability that any marble we draw is red is ? = 0.55 (independently, with replacement). We address the probability of getting no red marbles (? = 0) in the following cases:

4. We draw only one such sample. Compute the probability that $\nu = 0$. The closest answer is (closest is the answer that makes the expression —your answer? given option— closest to 0):

$P(\nu = 0) = (1 - 0.55)^{10} = 3.405 * 10^{-4}$

Answer: [b]

5. We draw only one such sample. Compute the probability that $\nu = 0$. The closest answer is (closest is the answer that makes the expression —your answer? given option— closest to 0):

$P(all\ \nu > 0) = (1 - P(\nu = 0))^{1000} = 0.71$
$P(at\ least\ one\ \nu = 0) = 1 - P(all\ \nu > 0) = 0.29$

Answer: [c]

Feasibility of Learning

Consider a boolean target function over a 3-dimensional input space $X = \{0, 1\}^3$ (instead of our ± 1 binary convention, we use $0, 1$ here since it is standard for boolean functions). We are given a data set D of five examples represented in the table below, where $y_n = f(x_n)$ for $n = 1, 2, 3, 4, 5$.

x_n			y_n
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1

Note that in this simple boolean case, we can enumerate the entire input space (since there are only $2^3 = 8$ distinct input vectors), and we can enumerate the set of all possible target functions (there are only $2^{2^3} = 256$ distinct boolean function on 3 boolean inputs).

Let us look at the problem of learning f. Since f is unknown except inside D, any function that agrees with D could conceivably be f. Since there are only 3 points in X outside D, there are only $2^3 = 8$ such functions.

The remaining points in X which are not in D are: 101, 110, and 111. We want to determine the hypothesis that agrees the most with the possible target functions. In order to quantify this, count how many of the 8 possible target functions agree with each hypothesis on all 3 points, how many agree with just 2 of the points, with just 1, and how many do not agree on any points. The final score for each hypothesis is computed as follows:

Score = (*# of target functions agreeing with hypothesis on all* 3 *points*) $*$ 3 + (*# of target functions agreeing with hypothesis on* 2 *points*) $*$ 2 + (*# of target functions agreeing with*

$hypothesis\ on\ 1\ point) * 1 + (\#\ of\ target\ functions\ agreeing\ with\ hypothesis\ on\ 0\ points) * 0$.

Eight possible target functions are:

f_1	0	0	0
f_2	0	0	1
f_3	0	1	0
f_4	0	1	1
f_5	1	0	0
f_6	1	0	1
f_7	1	1	0
f_8	1	1	1

[**a**] g returns 1 for all three points: score = 12

[**b**] g returns 0 for all three points: score = 12

[**c**] g is the XOR function, returns 0, 0, 1: score = 12

[**d**] g is the opposite of XOR, returns 1, 1, 0: score = 12

Answer: [**e**]

The Perceptron Learning Algorithm

My code for running the experiments to answer question number $7 - 10$ is uploaded here: `https://github.com/vkhoi/learning-from-data/blob/master/perceptron/perceptron.ipynb`

7. Answer: [**b**]

8. Answer: [**c**]

9. Answer: [**b**]

10. Answer: [**b**]