Aditya Joshi
15 th April 2015

CS 156 Machine Learning
Abu Yaser-Mostafa
Homework 1

Problem 1

The first choice does not involve learning as we know the target function already ignoring noise. There is nothing to learn.

The second case entails a data-set in which we have no labels for the data. The algorithm figures out decision boundaries based on techniques such as clustering. This is (ii) Unsupervised Learning.

The third case involves the agent to "learn" the system by making mistakes. This is called (iii) Reinforcement Learning in which a system essentially learns the parameters of a Markov decision process.

The answers are Not learning, Unsupervised Learning and Reinforcement Learning respectively.

Problem 2

Option (ii) is a typical Supervised learning problem. This therefore eliminates any option that does not contain (ii) that as an answer. Option (iv) could be a case of Reinforcement learning.

Problem 3

We decompose the problem as follows. We name the bags B_1 and B_2 with both black balls and a black and white ball respectively. We have observed a black ball b. We find the conditional probability of the bag being chosen being B_1 and B_2, and condition based on this event.

Now, the probability of the bag being B_1 is

$$P(B_1|b) = \frac{P(b|B_1) \cdot P(B_1)}{P(b)}$$

This value is $\frac{2}{3}$. Therefore, $P(B_2|b) = 1 - \frac{2}{3} = \frac{1}{3}$

Conditioning on these values, the probability that the next ball is black is

$$\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \boxed{\frac{2}{3}}$$

Problem 4

If we draw only one sample of 10 marbles, the probability of getting no red marbles is $(1 - 0.55)^{10} = \boxed{3 \times 10^{-4}}$

Problem 5

First let us consider a simpler problem, let us find the probability that $\nu \neq 0$ in **one** sample. This is simply $1 - P_{\text{single}}(\nu = 0)$ which is what we found out in the previous problem. Thus, for one sample, we have $P_{\text{single}}(\nu \neq 0) = 1 - (3 \times 10^{-4}) \approx .99965949$. Now, we want to find the probability that this happens for **all** 1000 samples. This is simply $P_{\text{all}}(\nu \neq 0)^{1000} = 0.7113688$. Finally to get $P_{\text{all}}(\nu = 0)$, we subtract $P_{\text{all}}(\nu \neq 0)$ from 1, giving us

$$1 - P_{\text{all}}(\nu \neq 0) \approx \boxed{0.288631}$$

Concisely, the computation done to find $P_{\text{all}}(\nu = 0)$ was $(1 - (1 - P_{\text{single}}(\nu = 0)^{1000}))$