Machine Learning, 2021 Spring
Homework 1 and Solution

Due on 12:59 MAR 15, 2021

We pick a random sample of N independent marbles (with replacement) from this bin (μ is the probability of red marbles), and observe the fraction ν of red marbles within the sample (Figure 1).

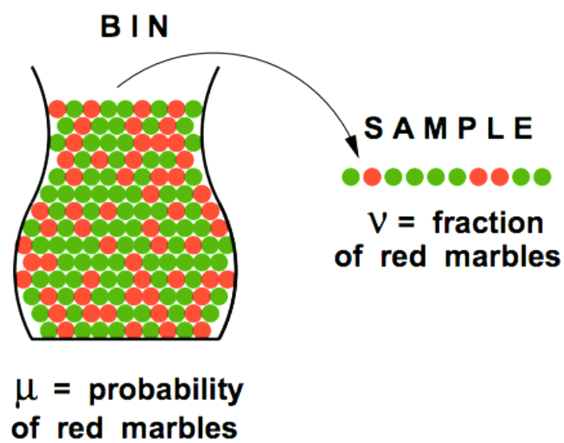

Figure 1

Problem 1

If $\mu = 0.9$, what is the probability that a sample of 10 marbles will have $\nu \leq 0.1$? [Hints: 1. Use binomial distribution. 2. The answer is a very small number.] [2pts]

Solution Let X be the number of red marbles we take from the bin, clearly, we have

$$\mathbb{P}(X = k) = \binom{10}{k} \mu^k (1-\mu)^{10-k}, k = 0, \dots, 10 \tag{1}$$

Evaluates $\nu \leq 0.1$ leads to

$$\nu \leq 0.1 \iff \frac{X}{10} \leq 0.1 \iff X \leq 1 \iff X = 0, 1 \tag{2}$$

thus,

$$\mathbb{P}(\nu \leq 0.1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = 0.1^{10} + 10 * 0.9 * 0.1^9 = 9.1 * 10^{-9} \tag{3}$$

Problem 2

If $\mu = 0.9$, use the Hoeffding Inequality to bound the probability that a sample of 10 marbles will have $\nu \le 0.1$. [2pts]

Solution Since $\nu \le 0.1$, we have $|\mu - \nu| \ge 0.8$, by Hoeffding Inequality with $\epsilon = 0.8$ and $N = 10$:

$$\mathbb{P}\left[|\mu - \nu| \ge 0.8\right] \le 2e^{-2*0.8^2*10} = 2.76 * 10^{-6} \tag{4}$$

The result in **Problem 1** justifies our answer.

Problem 3

We are given a data set $\mathcal{D}$ and of 25 training examples from an unknown target function $f : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \{-1, +1\}$. To learn f, we use a simple hypothesis set $\mathcal{H} = \{h_1, h_2\}$ and, where h_1 is the constant $+1$ function and h_2 is the constant -1.

We consider two learning algorithms, *S* (smart) and *C* (crazy). *S* chooses the hypothesis that agrees the most with $\mathcal{D}$ and *C* chooses the other hypothesis deliberately. Let us see how these algorithms perform out of sample from the deterministic and probabilistic points of view. Assume in the probabilistic view that there is a probability distribution on $\mathcal{X}$, and let $\mathbb{P}[f(x) = +1] = p$.

(a) Can *S* produce a hypothesis that is guaranteed to perform better than random on any point outside $\mathcal{D}$? [1pt]

(b) Assume for the rest of the exercise that all the examples in $\mathcal{D}$ have $y_n = +1$. Is it possible that the *C* hypothesis that produces turns out to be better than the hypothesis that *S* produces? [1pt]

(c) If $p = 0.9$, what is the probability that *S* will produce a better hypothesis than *C*? [2pts]

(d) Is there any value of p for which it is more likely that *C* will produce a better hypothesis than *S*? [2pts]

Solution

(a) No, since $\mathbb{P}[f(x) = +1] = p$, if $p < 0.5$, then it's more likely we pick a negative sample ($f(x) = -1$), which means *C* is doing better than *S*.

(b) If the data are sampled *i.i.d.*, given the information about the training dataset, we can estimate p with Hoeffding inequality, clearly, $\nu = 1$, we can estimate p with different ϵ:

$$\mathbb{P}\left[|p - 1| > 0.2\right] \le 2e^{-2*0.2^2*25} = 0.27 \Leftrightarrow \mathbb{P}(p < 0.8) \le 0.27 \tag{5}$$

$$\mathbb{P}\left[|p - 1| > 0.3\right] \le 2e^{-2*0.3^2*25} = 0.022 \Leftrightarrow \mathbb{P}(p < 0.7) \le 0.022 \tag{6}$$

$$\mathbb{P}\left[|p - 1| > 0.4\right] \le 2e^{-2*0.4^2*25} = 6.7 * 10^{-4} \Leftrightarrow \mathbb{P}(p < 0.6) \le 6.7 * 10^{-4} \tag{7}$$

$$\mathbb{P}\left[|p - 1| > 0.5\right] \le 2e^{-2*0.5^2*25} = 7.45 * 10^{-6} \Leftrightarrow \mathbb{P}(p < 0.5) \le 7.45 * 10^{-6} \tag{8}$$

as we can see from above, it almost impossible that $p < 0.5$! So we may believe that *S* performs better than *C* outside of the dataset under a low error rate.[a]

Now if the data are not sampled *i.i.d.*, we are not able to generalize the information we obtained from the training dataset, so *C* may still performs better than *S*.

(c) Since all the examples in $\mathcal{D}$ have $y_n = +1$, h_1 agrees most with $\mathcal{D}$, thus *S* will choose h_1. *S* produces a better hypothesis than *C* means that it has smaller out-of-sample error:

$$\begin{aligned}
\mathbb{P}[E_{out}(h_1) < E_{out}(h_2)] &= \mathbb{P}\left[\mathbb{P}\left[f(x) \ne h_1(x)\right] < \mathbb{P}\left[f(x) \ne h_2(x)\right]\right] && (9)\\
&= \mathbb{P}\left[\mathbb{P}\left[f(x) = -1\right] < \mathbb{P}\left[f(x) = +1\right]\right] && (10)\\
&= \mathbb{P}\left[1 - p < p\right] = \mathbb{P}\left[0.1 < 0.9\right] = 1 && (11)
\end{aligned}$$

(d) Similar to (c), we have

$$\begin{aligned}
\mathbb{P}[E_{out}(h_1) > E_{out}(h_2)] &= \mathbb{P}\left[\mathbb{P}\left[f(x) \neq h_1(x)\right] > \mathbb{P}\left[f(x) \neq h_2(x)\right]\right] && (12)\\
&= \mathbb{P}\left[\mathbb{P}\left[f(x) = -1\right] > \mathbb{P}\left[f(x) = +1\right]\right] && (13)\\
&= \mathbb{P}\left[1 - p > p\right] = \begin{cases} 1, p < 0.5 \\ 0, p \leq 0.5 \end{cases} && (14)
\end{aligned}$$

C produces a better result means that $\mathbb{P}[E_{out}(h_1) > E_{out}(h_2)] > 0.5$, from the above result it means $p < 0.5$.

[a] For more formal description about error rate, please refer to hypothesis test.