Machine Learning, 2021 Spring
Homework 1

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 ??).

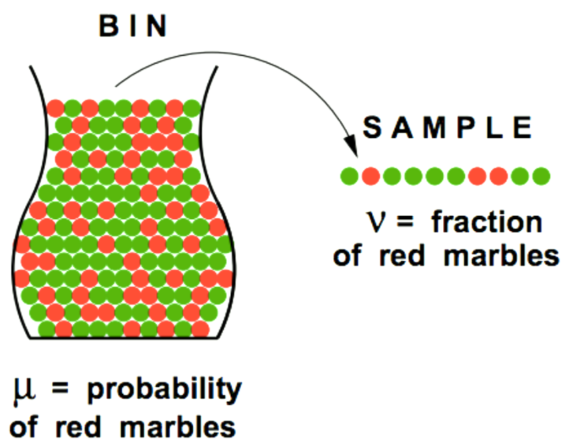


Figure 1

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]

1. Let X be the color of one marble, $X \sim Bern(0.9)$

$X_1, \ldots, X_{10}$ are ten i.i.d samples, $Y = X_1 + \cdots + X_n \sim Bin(10, 0.9)$

$Pr(\nu \leq 0.1) = Pr(Y \leq 1)$

$= Pr(Y=0) + Pr(Y=1)$

$= (0.1)^{10} + \binom{10}{1}(0.1)^9(0.9)^1$

$= 9.1 \times 10^{-9}$

Problem 2

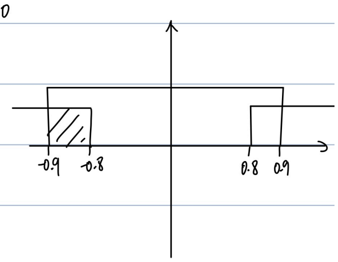
If $\mu = 0.9$, use the Hoeffding Inequality to bound the probability that a sample of 10 marbles will have $\nu \leq 0.1$. [2pts]

2. Hoeffding inequality: $Pr(|\nu-\mu|>\varepsilon) \leq 2e^{-2\varepsilon^2 m}$, $\varepsilon>0$

$Pr(|\nu-\mu| \leq \varepsilon) \geq 1-2e^{-2\varepsilon^2 m}$

$Pr(\nu \leq 0.1) = Pr(-0.9 \leq \nu-\mu \leq -0.8)$

$= \frac{1}{2}[Pr(|\nu-\mu|>0.8) + Pr(|\nu-\mu| \leq 0.9) - 1]$



As $Pr(|\nu-\mu|>0.8) \leq 2e^{-2\times(0.8)^2\times 10}$

$Pr(|\nu-\mu| \leq 0.9) \leq 1$

$\leq \frac{1}{2} \cdot 2e^{-2\times(0.8)^2\times 10}$

$= 2.76\times 10^{-6}$

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]

3.(a) No eg. $\forall(x,y)\in D: y=+1$, $\forall(x,y)\in \mathcal{X}/D, y=-1$

then S will choose h_1, C will chose h_2, h_2 will always perform better than h_1(chose by S) on any point outside D

(b). Yes, consider $|\mathcal{X}|=100$, $|D|=25$, $\forall(x,y)\in D: y=+1$, $\forall(x,y)\in \mathcal{X}/D, y=-1$

S will produce h_1 based on D, and C will produce h_2, on $\mathcal{X}$ h_1(produced by C) will be better than h_2(produced by S)

(C) As $\forall(x,y)\in D: y=+1$

$\therefore$ S will produce h_1

$Pr(h_1(x)=f(x))=Pr(+1=f(x))=0.9$

$Pr(h_2(x)=f(x))=Pr(-1=f(x))=0.1$

$\therefore Pr(S \text{ better than } C)=Pr(Pr(h_1(x)=f(x))>Pr(h_2(x)=f(x))$

$=1$

(d) $Pr(h_1(x)=f(x))=Pr(+1=f(x))=p$ [S always produce h_1]

$Pr(h_2(x)=f(x))=Pr(-1=f(x))=1-p$

C better than S $\Leftarrow 1-p>p$

$\therefore \forall p\in[0,\frac{1}{2})$, C produce better than S