Machine Learning Homework #6

R01922024 Qing-Cheng Li

1

對每一個維度來看，都可以切出R-L+2個θ，以及s={+1,-1}，一共有2*(R-L+2)種組合，但其中頭尾有一組分法是重複的，所以每個維度實際上能產生的g一共是2*(R-L+1)個。若有d個維度，則最多可有d*2*(R-L+1)個g。

以d=2,(L,R)=(1,6)，則有24個g：
$g_{1,1,0.5}(x), g_{-1,1,0.5}(x), g_{1,1,1.5}(x), g_{-1,1,1.5}(x), g_{1,1,2.5}(x), g_{-1,1,2.5}(x), g_{1,1,3.5}(x), g_{-1,1,3.5}(x), g_{1,1,4.5}(x),$
$g_{-1,1,4.5}(x), g_{1,1,5.5}(x), g_{-1,1,5.5}(x), g_{1,2,0.5}(x), g_{-1,2,0.5}(x), g_{1,2,1.5}(x), g_{-1,2,1.5}(x), g_{1,2,2.5}(x), g_{-1,2,2.5}(x),$
$g_{1,2,3.5}(x), g_{-1,2,3.5}(x), g_{1,2,4.5}(x), g_{-1,2,4.5}(x), g_{1,2,5.5}(x), g_{-1,2,5.5}(x)$

2

對每個維度d而言，總共會有2(R-L+1)個g，這其中會有$2|x'_d - x_d|$個g是把x’和x分成不同類，剩下$2(R-L+1) - 2|x'_d - x_d|$ 個是同類，所以對維度d算sum(g*g)，會有$2(R-L+1) - 2|x'_d - x_d|$ 個+1，$2|x'_d - x_d|$個-1，總和為$2(R-L+1) - 4|x'_d - x_d|$ ，故K(x,x’) =
$$\sum_{d=1}^{D} 2(R-L+1) - 4|x'_d - x_d|$$

3

已知如果$y_n g_t(x_n) = 1$ 代表正確，要除$\lozenge_t$，反之則乘以$\lozenge_t$。可以利用$e^{ln(x)} = x$ 的特性，把這兩種情況都寫成乘法：乘以$e^{-y_n g_t(x_n) ln(\lozenge_t)}$。如果$y_n g_t(x_n) = 1$ ，乘以$e^{-y_n g_t(x_n) ln(\lozenge_t)} = e^{-ln(\lozenge_t)} = \frac{1}{\lozenge_t}$ ，若$y_n g_t(x_n) = -1$ ，乘以$e^{-y_n g_t(x_n) ln(\lozenge_t)} = e^{ln(\lozenge_t)} = \lozenge_t$ 。又，$\alpha_t = ln(\lozenge_t)$ 。

初始$u_n^{(1)} = \frac{1}{N}$ ，$u_n^{(t+1)} = u_n^{(1)} \times e^{-y_n \alpha_1 g_1(x_n)} \times \cdots \times e^{-y_n \alpha_t g_t(x_n)} = \frac{1}{N} e^{-y_n \sum_{\tau=1}^{t} \alpha_\tau g_\tau(x_n)}$ ，
$U^{(t+1)} = \sum_{n=1}^{N} u_n^{t+1} = \frac{1}{N} \sum_{n=1}^{N} e^{-y_n \sum_{\tau=1}^{t} \alpha_\tau g_\tau(x_n)}$ 。

4

$E_{in}(G) = \frac{\sum_{n=1}^{N} [[y_n \neq G(x_n)]]}{N}$, for $x_n, [[y_n \neq G(x_n)]] \leq exp(-y_n G(x_n))$,
$$E_{in}(G) = \frac{1}{N} \sum_{n=1}^{N} [[y_n \neq G(x_n)]] \leq \frac{1}{N} \sum_{n=1}^{N} exp(-y_n G(x_n)) = \frac{1}{N} \sum_{n=1}^{N} exp(-y_n \sum_{\tau=1}^{T} \alpha_\tau g_\tau(x_n)) = U^{(T+1)}$$

5

$$U^{(t+1)} = (\sum_{n=1}^{N} u_n^{(t)})[(\epsilon_t)(\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}) + (1-\epsilon_t)(\sqrt{\frac{\epsilon_t}{1-\epsilon_t}})] = U^{(t)}(\sqrt{\epsilon_t(1-\epsilon_t)} + \sqrt{\epsilon_t(1-\epsilon_t)})$$
$$= U^{(t)} 2\sqrt{\epsilon_t(1-\epsilon_t)}$$

6

$$\sqrt{\epsilon_t(1-\epsilon_t)} \le \sqrt{\epsilon(1-\epsilon)} \Leftrightarrow \epsilon_t(1-\epsilon_t) \le \epsilon(1-\epsilon) \Leftrightarrow \epsilon(1-\epsilon) - \epsilon_t(1-\epsilon_t) \ge 0$$

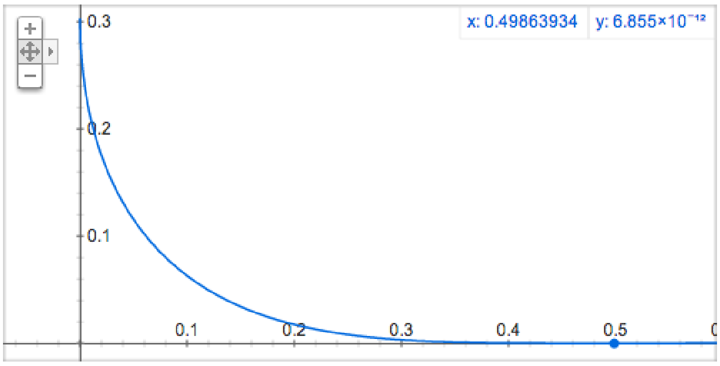
$$\epsilon(1-\epsilon) - \epsilon_t(1-\epsilon_t) = \epsilon - \epsilon_t + \epsilon_t^2 - \epsilon^2 = (\epsilon_t - \epsilon)(\epsilon_t + \epsilon) - (\epsilon_t - \epsilon) = (\epsilon_t - \epsilon)(\epsilon_t + \epsilon - 1) > 0$$

So, $\sqrt{\epsilon_t(1-\epsilon_t)} \le \sqrt{\epsilon(1-\epsilon)}$ is True.

7.

When $\epsilon < \frac{1}{2}$, it' s True.

0.5*exp((-2)*(0.5-x)*(0.5-x))-sqrt(x*(1-x)) 的圖表



8.

$$U^{(t+1)} = (2\sqrt{\epsilon_1(1-\epsilon_1)}\,) \times \cdots \qquad \text{(p5,p6)}$$
$$\times (2\sqrt{\epsilon_t(1-\epsilon_t)}\,) \le (2\sqrt{\epsilon(1-\epsilon)}\,)^T$$

$$\sqrt{\epsilon(1-\epsilon)} \le \frac{1}{2}\, exp(-2(\tfrac{1}{2}-\epsilon)^2) \qquad \text{(p7)}$$
$$\Leftrightarrow 2\sqrt{\epsilon(1-\epsilon)} \le exp(-2(\tfrac{1}{2}-\epsilon)^2)$$
$$\Leftrightarrow (2\sqrt{\epsilon(1-\epsilon)}\,)^T \le exp(-2T(\tfrac{1}{2}-\epsilon)^2)$$
$$\Rightarrow U^{(t+1)} \le exp(-2T(\tfrac{1}{2}-\epsilon)^2)$$

9.

令$T = k \times log(N)$

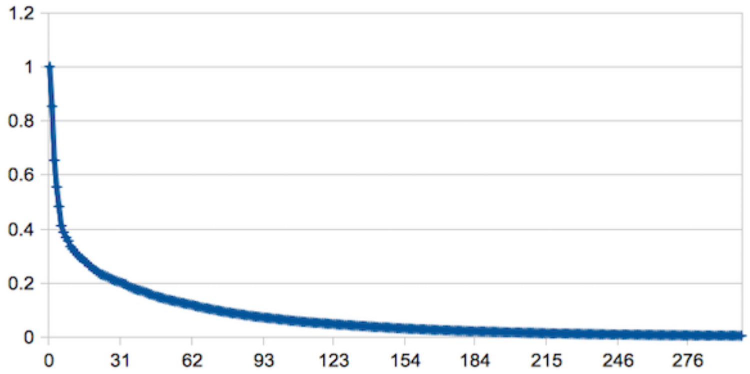
$$E_{in}(G) \le U^{T+1} \le exp(-2klog(N)(\tfrac{1}{2}-\epsilon)^2)$$
$$= (\tfrac{1}{N})^{2k(\frac{1}{2}-\epsilon)^2}$$

當k是一個夠大的常數時，$E_{in}(G)$ 將會夠接近0，便可當作等於0了。

10,11.

```
$ python adaboost.py
```

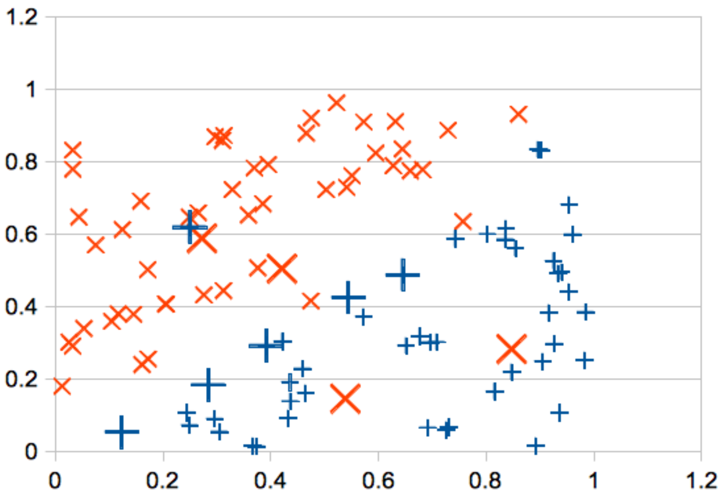
12.



13.

確實觀察到了$U^t \geq E_{in}(G_t)$ 的現象。也有觀察到當T越來越大時，$\geq E_{in}(G_t)$ 基本上等於0的現象。另外$U^{(t+1)}$確實等於$U^{(t)} 2\sqrt{\epsilon_t(1-\epsilon_t)}$ 。

14.



較大的u出現在一些看來應該是error（明顯與周遭不同類）或是邊界上。

15.

```
$ python cart.py
```

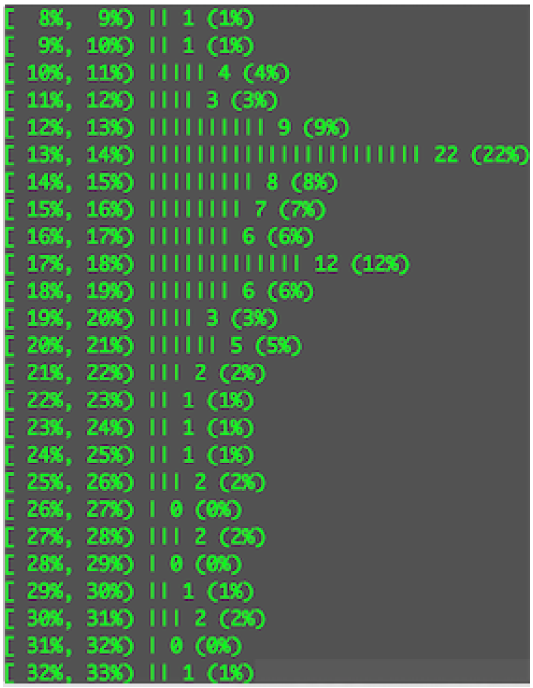
It is too big, please see the result of running cart.py.

16.

Ein(G) = 0, Eout(G) = 0.141

17.

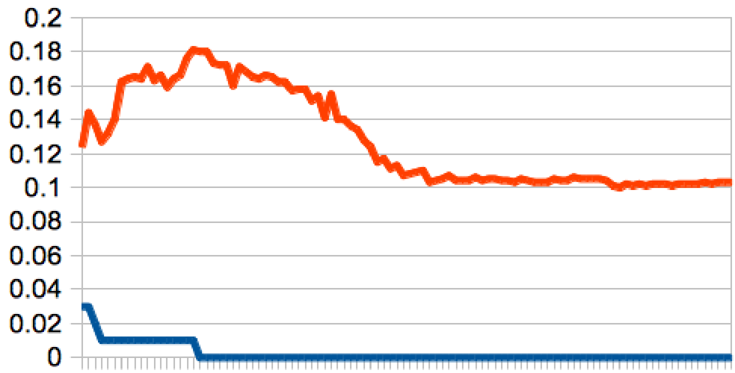
$ python randomForset.py



```
[  8%,  9%) || 1 (1%)
[  9%, 10%) || 1 (1%)
[ 10%, 11%) ||||| 4 (4%)
[ 11%, 12%) |||| 3 (3%)
[ 12%, 13%) |||||||||| 9 (9%)
[ 13%, 14%) ||||||||||||||||||||||| 22 (22%)
[ 14%, 15%) ||||||||| 8 (8%)
[ 15%, 16%) |||||||| 7 (7%)
[ 16%, 17%) ||||||| 6 (6%)
[ 17%, 18%) ||||||||||||| 12 (12%)
[ 18%, 19%) ||||||| 6 (6%)
[ 19%, 20%) |||| 3 (3%)
[ 20%, 21%) |||||| 5 (5%)
[ 21%, 22%) ||| 2 (2%)
[ 22%, 23%) || 1 (1%)
[ 23%, 24%) || 1 (1%)
[ 24%, 25%) || 1 (1%)
[ 25%, 26%) ||| 2 (2%)
[ 26%, 27%) | 0 (0%)
[ 27%, 28%) ||| 2 (2%)
[ 28%, 29%) | 0 (0%)
[ 29%, 30%) || 1 (1%)
[ 30%, 31%) ||| 2 (2%)
[ 31%, 32%) | 0 (0%)
[ 32%, 33%) || 1 (1%)
```

看似為一個常態分布，平均蠻接近前題之Eout。

18.



隨著森林裡面的樹越來越多， Ein最終能夠一如單顆完整的決策樹將至0，同樣隨著森林裡頭樹木的增長，Eout可以下降到比只使用一個完整的決策樹還要低，但最終仍有瓶頸不會一路降低至0。

19.

PLA，因為這個部分讓我很明確的理解了「機器」「學習」這件事情是如何發生的，在我心中種下了一個「機器」**真的**有能力「學習」的種子。

20.

AdaBoost，因為單看投影片我還是沒有辦法理解，必須搭配上課的口說內容，這樣想來是有某些東西投影片上沒有寫的太直白。