HW6

1.

$\Phi_{\mathrm{ds}}(x)^T \phi_{\mathrm{ds}}(x')$這個東西展開就是$\sum_{i=1}^{d} \sum_{\theta=L+0.5}^{R-0.5} sign(x_i - \theta) sign(x_i' - \theta)$，觀察可以發現當$sign(x_i - \theta)$跟$sign(x_i' - \theta)$同號時$sign(x_i - \theta) sign(x_i' - \theta)$是正的，其他時是負的。只有當$\theta$在$x_i$跟$x_i'$之間時$sign(x_i - \theta)$跟$sign(x_i' - \theta)$會不同號，而且因為 s 有 2 種，所以對於每個 i 而言，$sign(x_i - \theta)$跟$sign(x_i' - \theta)$不同號會發生 2$\|x_i - x_i'\|$次。所以對於 x 跟 x'而言，如果$sign(x_i - \theta)$跟$sign(x_i' - \theta)$全部都是同號，因為 s 有 2 種，R 跟 L 之間有 R-L 個θ，有 d 種 i，所以$\sum_{i=1}^{d} \sum_{\theta=L+0.5}^{R-0.5} sign(x_i - \theta) sign(x_i' - \theta) = 2d(R - L)$。但現在其中有 2$\|x_i - x_i'\|$是異號的，所以$\Phi_{\mathrm{ds}}(x)^T \phi_{\mathrm{ds}}(x') = 2d(R-L) - 2\|x_i - x_i'\| - 2\|x_i - x_i'\| = 2d(R-L) - 4\|x_i - x_i'\|$。

2.

對於所有α_{n}，定$\mathrm{A_n} = \frac{\alpha_{\mathrm{n}}}{u}$。顯然當 C 變成 C/u 時，$\mathrm{A_n}$是符合 0 到 C/u 之間的限制的。原本的 dual problem 要 min 的公式是$\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum_{n=1}^{N} \alpha_n$。當 K=uK+v 時，公式變成$\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m (uK(x_n, x_m) + v) - \sum_{n=1}^{N} \alpha_n = \frac{\mathrm{u}}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} A_n A_m y_n y_m K(x_n, x_m) - u\sum_{n=1}^{N} A_n + \mathrm{v}(\sum_{n=1}^{N}\sum_{m=1}^{N} A_n A_m y_n y_m) = \frac{\mathrm{u}}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} A_n A_m y_n y_m K(x_n, x_m) -$

$u\sum_{n=1}^{N} A_n + \mathrm{v}(\sum_{n=1}^{N} A_n y_n)(\sum_{m=1}^{N} A_m y_m)$，因為 dual problem 有個 constraint 是$\sum_{n=1}^{N} A_n y_n = 0$，所以

$\frac{\mathrm{u}}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} A_n A_m y_n y_m K(x_n, x_m) - u\sum_{n=1}^{N} A_n + \mathrm{v}(\sum_{n=1}^{N} A_n y_n)(\sum_{m=1}^{N} A_m y_m) = \frac{\mathrm{u}}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} A_n A_m y_n y_m K(x_n, x_m) - u\sum_{n=1}^{N} A_n$，也就是原本公式乘上 u 倍。因此解完 QP 後得到的 A 就是原本α解的 1/u 倍。而最終 gsvm 的公式是$\mathrm{g_{svm}}(x) = sign(\sum_{sv} \alpha_n y_n K(x_n, x) + b)$，當 K=uK+v，變成 $sign(\sum_{sv} A_n y_n (uK(x_n, x) + v) + b) = sign(\sum_{sv} \alpha_n y_n K(x_n, x) + b + v\sum_{sv} \alpha_n y_n) = \ sign(\sum_{sv} \alpha_n y_n K(x_n, x) + b)$，因為$\sum_{n=1}^{N} \alpha_n y_n = 0$這個 constraint。所以解還是跟原本一樣。

3

講義裡有個公式$\mathrm{avg}\big(\mathrm{E_{out}}(\mathrm{g_t})\big) = \mathrm{avg}(\varepsilon(\mathrm{g_t} - G)^2) + \mathrm{E_{out}}(G)$，所以$1 \geq \frac{\mathrm{E_{out}}(G)}{\mathrm{avg}(\mathrm{E_{out}}(\mathrm{g_t}))} = \frac{\mathrm{E_{out}}(G)}{\frac{\mathrm{E}}{17}} \rightarrow \frac{\mathrm{E_{out}}(G)}{\mathrm{E}} \leq \frac{1}{17}$

4

每次拿到特定 data 的機率是 1-1/N，做 3/4*N 次，所以當 N 趨近無限大時，$\left(1 - \frac{1}{\mathrm{N}}\right)^{\frac{3}{4}\mathrm{N}} = \left(\frac{\mathrm{N}-1}{\mathrm{N}}\right)^{\frac{3}{4}\mathrm{N}} = \frac{1}{\left(\frac{\mathrm{N}}{\mathrm{N}-1}\right)^{\frac{3}{4}\mathrm{N}}} = \frac{1}{\left(1+\frac{1}{\mathrm{N}-1}\right)^{\frac{3}{4}\mathrm{N}}} = \frac{1}{\mathrm{e}^{\frac{3}{4}}}$。

5

Adaboost 的算法可以確保不管前面正確率多高，只要對 u 做出更新，當下這輪的 u 一定保證$\frac{\sum_{n:y_n>0} u_n}{\sum_{n:y_n<0} u_n} = 1$。所以答案是 1。

6

$\mathrm{U}_1 = 1$，對於U_t而言，ε_t代表有$\varepsilon_t \mathrm{U}_\mathrm{t}$這麼多的 u 是屬於 incorrect，需要乘上$\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}$。其他部分是除以$\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}}$，，所以$\mathrm{U}_{\mathrm{t}+1} = \varepsilon_t \mathrm{U}_\mathrm{t} * \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} + (1-\varepsilon_t)\mathrm{U}_\mathrm{t} * \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}\mathrm{U}_\mathrm{t}$，所以$\frac{\mathrm{U}_{\mathrm{t}+1}}{\mathrm{U}_\mathrm{t}} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$。因此$\frac{\mathrm{U}_{\mathrm{T}+1}}{\mathrm{U}_1} = 2^\mathrm{T} \prod_{t=1}^{T} \sqrt{\varepsilon_t(1-\varepsilon_t)}$

7.

因為α_t就是 minimize $\sum_{n=1}^{N}\left((y_n - s_n) - \eta g_t(x_n)\right)^2$的解，所以$\sum_{n=1}^{N}\left((y_n - s_n) - \eta g_t(x_n)\right)^2$對$\eta$的微分=0。而$\sum_{n=1}^{N}\left((y_n - s_n) - \eta g_t(x_n)\right)^2$微分是$\sum_{n=1}^{N} -2\left((y_n - s_n) - \eta g_t(x_n)\right) g_t(x_n) = 0$。解是$\eta = \alpha_\mathrm{t}$。所以$\sum_{n=1}^{N}\left((y_n - s_n) - \alpha_\mathrm{t} g_t(x_n)\right) g_t(x_n) = 0$。用公式$\mathrm{s_n} = s_n + \alpha_\mathrm{t} g_t(x_n)$把$\mathrm{s_n}$更新成最後的$\mathrm{s_n}$後，就可以得到$\sum_{n=1}^{N}\left((y_n - s_n)\right) g_t(x_n) = 0$

8.

講義上 output layer 的微分沒有 w 的因素，不受影響。但是對於

其他點的微分，$\frac{\partial \mathrm{e_n}}{\partial w_{ij}^{l}} = x_i^{l-1}\delta_j^l$，講義中，$\delta_j^l$的算式有包含 l+1 層的 w，所以除了 output layer 外，其他 layer 的 gradient 計算結果都是 0。但因為初始所有 w 都是 0，所以對於 output layer 的x_i^{l-1}也都是 0。所以 output layer 的 gradient 也是 0。所以所有 w 的 gradient 都是 0。

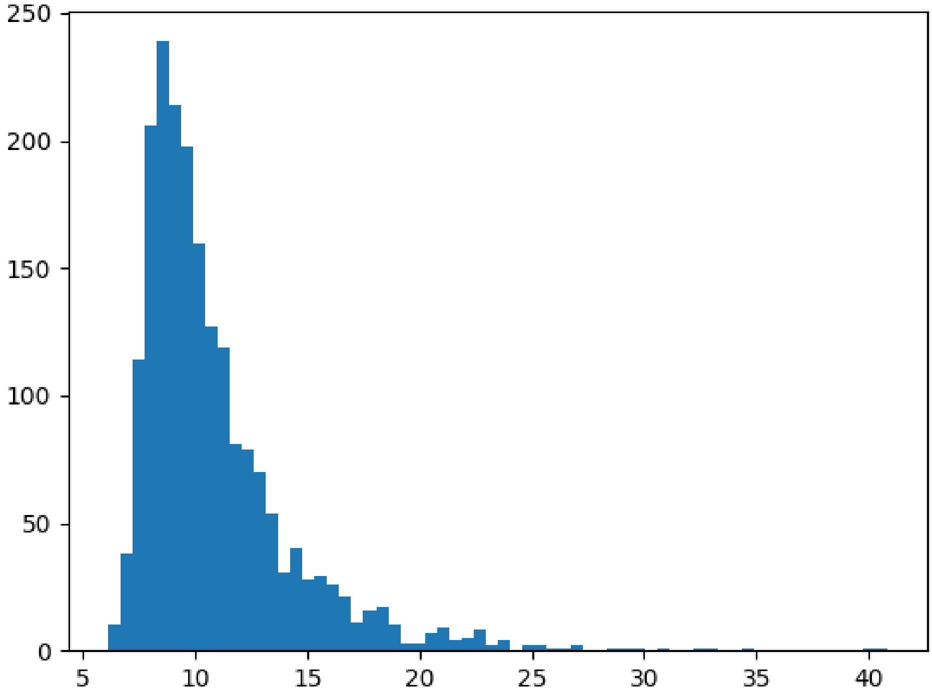
9.

我設最大深度是 5。要不然跑太慢，後面的題目跑超久。

我的 Eout 是 9.082906857727737

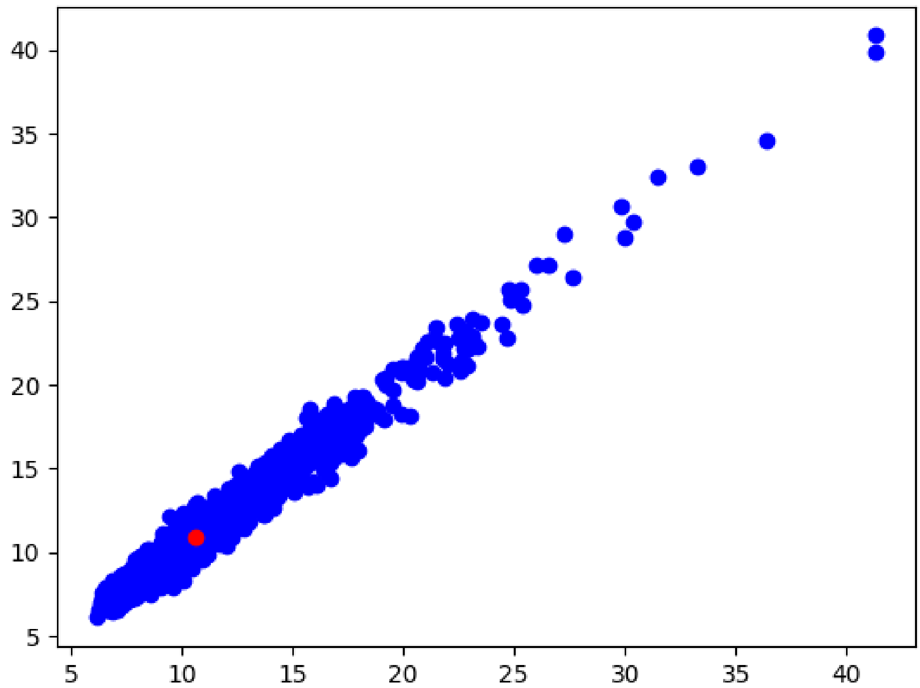
10.

這是我畫的圖



11

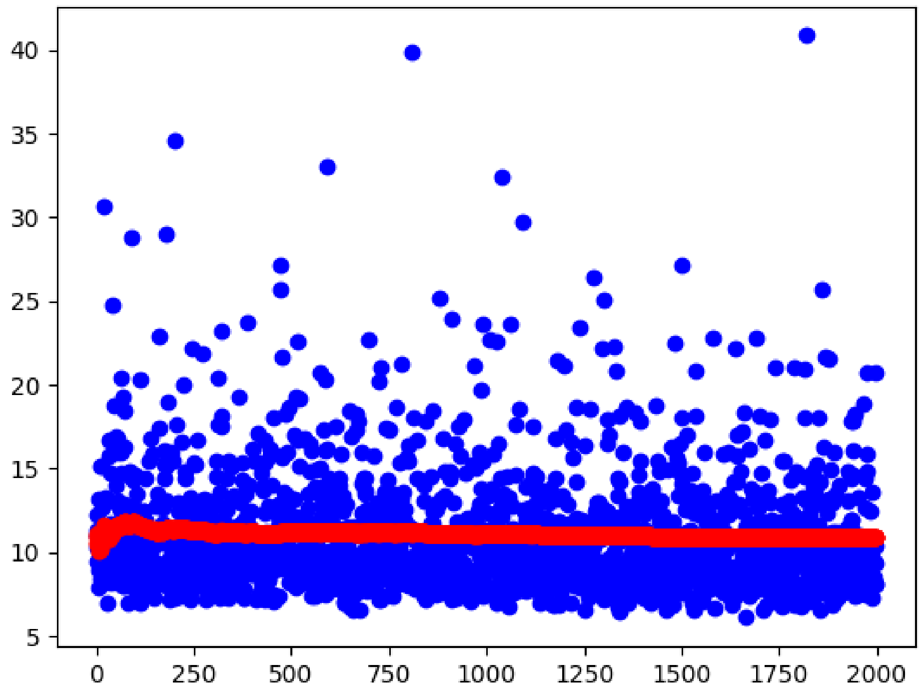
這是我畫的圖



可以發現說其實 G 比大部分的 Eout 還要好。所以選 G 是個比較穩定的選擇。

12

這是我畫的圖



其實 Eout 晃動幅度非常大，很不穩定。但 G 相對穩定在下降。

明顯穩定很多。