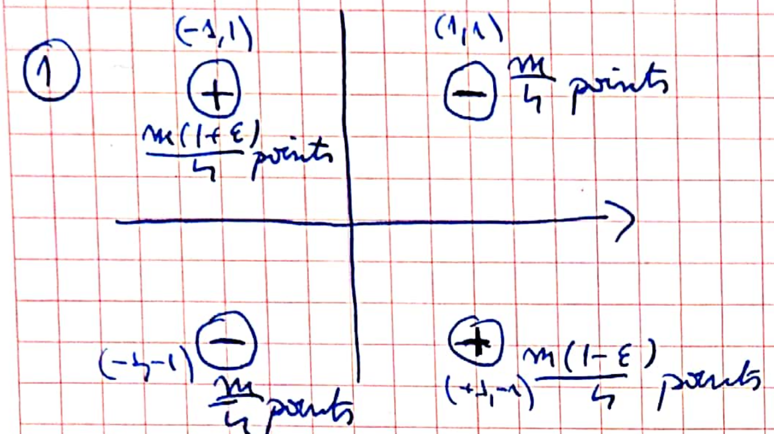
Seminar class 6

①



$\frac{m(1+\varepsilon)}{4} + \frac{m(1-\varepsilon)}{4} = \frac{m}{2}$ points with + label

$\frac{m}{4} + \frac{m}{4} = \frac{m}{2}$ points with − label

The probability distribution of the training point (−1, 1) with label + is $\frac{\frac{m(1+\varepsilon)}{4}}{m} = \frac{1+\varepsilon}{4}$. For point (1, −1) we obtain $\frac{1-\varepsilon}{4}$, for points (1, 1) and (−1, −1) with label − we obtain $\frac{1}{4}$.

The initial problem with m points in the training sample is similar with the problem with 4 points with the corresponding probabilities.

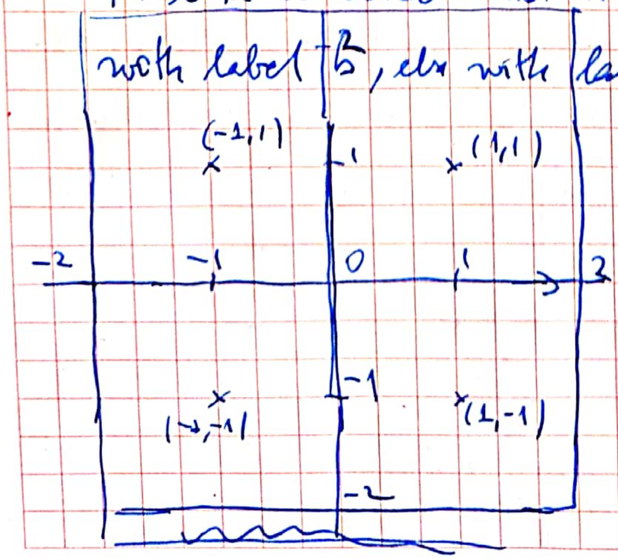
$$S = \{(\underset{\text{point}}{\underset{\downarrow}{(-1,1)}}, \underset{\text{label}}{\underset{\updownarrow}{+1}}), ((+1,-1), +1), ((+1,+1), -1), ((-1,-1), -1)\}$$

$$D^{(1)} : \begin{pmatrix} (-1,1) & (1,-1) & (1,1) & (-1,-1) \\ \frac{1+\varepsilon}{4} & \frac{1-\varepsilon}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Base hypothesis class = decision stumps in $\mathbb{R}^2$

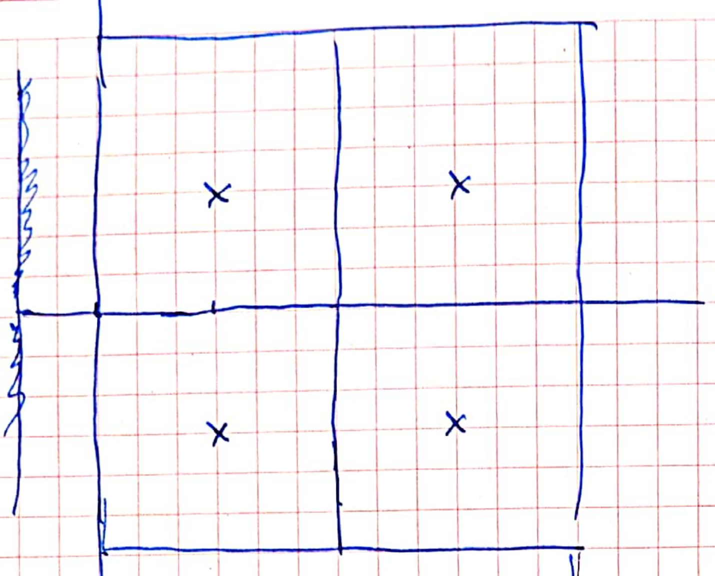
$$\mathcal{H}^2_{DS} = \{h_{i,\theta,b} : \mathbb{R}^2 \to \{-1,1\}, h_{i,\theta,b}(x_1,x_2) = \text{sign}(\theta - x_i)\cdot b, \ 1 \le i \le 2, \ \theta \in \mathbb{R}, \ b \in \{+1,-1\}\}$$

= pick a coordinate i (1 or 2), project the input $x = (x_1, x_2)$ on the i-th coordinate and obtain x_i; if $x_i \le \theta$ label the example x_i with label +b, else with label −b.



For our problem we can see that we can take set of representative thresholds θ for be θ = {−2, 0, 2}.

So we have at most 12 base classifiers $h_{1,-2,1}; h_{1,-2,-1}$ —— ; $h_{2,2,-1}$



+ | −

$h_{1,-2,+1}$ → project on x_1, compare to −2, all points < −2 get label 1, all other get label −1

$h_{1,-2,-1}$ – project on x_1, compare to −2, all point < −2 get label −1, all other get label +1

− | +

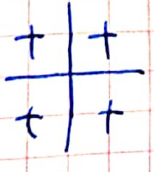
+ | −

$h_{1,+2,+1}$

project on x_1, compare to 2, all points < 2 get a label +1, all other get label −1

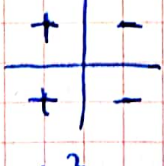
So we see that on our training set $h_{1,-2,-1}$ and $h_{1,+2,+1}$ will have the same behavior (all points will receive label +1).

If we analyze the behavior of all 12 base classifiers (decision stumps in $\mathbb{R}^2$) we will see that in the end there are only 6 unique base classifiers.

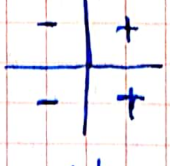


h^1

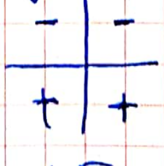
h^2



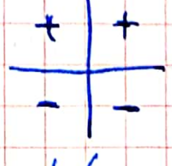
h^3



h^4



h^5



h^6

So we have $B = \{h^1, h^2, h^3, h^4, h^5, h^6\}$

Round 1

- distribution $\mathcal{D}^{(1)}: \left(\begin{array}{cccc} (-1,1) & (1,-1) & (1,1) & (-1,-1) \\ \frac{1+\varepsilon}{4} & \frac{1-\varepsilon}{4} & \frac{1}{4} & \frac{1}{4} \end{array}\right)$

- select the best classifier from $\mathcal{H}$, the one with minimum empirical risk

$$L_{\mathcal{D}^{(1)}}(h^1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

$$L_{\mathcal{D}^{(1)}}(h^2) = \frac{1+\varepsilon}{4} + \frac{1-\varepsilon}{4} = \frac{1}{2}$$

$$L_{\mathcal{D}^{(1)}}(h^3) = \frac{1}{4} + \frac{1-\varepsilon}{4} = \frac{1}{2} - \frac{\varepsilon}{4}$$

$$L_{\mathcal{D}^{(1)}}(h^4) = \frac{1+\varepsilon}{4} + \frac{1}{4} = \frac{1}{2} + \frac{\varepsilon}{4}$$

$$L_{\mathcal{D}^{(1)}}(h^5) = \frac{1+\varepsilon}{4} + \frac{1}{4} = \frac{1}{2} + \frac{\varepsilon}{4}$$

$$L_{\mathcal{D}^{(1)}}(h^6) = \frac{1}{4} + \frac{1-\varepsilon}{4} = \frac{1}{2} - \frac{\varepsilon}{4}.$$

So, the minimum achievable error is $\frac{1}{2} - \frac{\varepsilon}{4}$ and it is attained by two classifiers h^3 and h^6. Let's choose h^3 ~~as~~ our ~~the~~ weak classifier

$$h^3 = h_{1,0,+1}.$$

So, for $t = 1$ (round 1) we have $h_t = h^3 = h_{1,0,-1}$.

The error of the base classifier is $\varepsilon_1 = \frac{1}{2} - \frac{\varepsilon}{4}$.

$$w_1 = \frac{1}{2}\ln\left(\frac{1}{\varepsilon_1} - 1\right) = \frac{1}{2}\left(\ln\left(\frac{4}{2-\varepsilon} - 1\right) = \ln\left(\frac{2+\varepsilon}{2-\varepsilon}\right)^{\frac{1}{2}} = \ln\sqrt{\frac{2+\varepsilon}{2-\varepsilon}}.$$

Based on $\mathcal{D}^{(1)}$ we will build $\mathcal{D}^{(2)}$. Examples correctly classified at round 1 will have now the weight decreased, examples misclassified at round 1 will have their weight increased.

$$\mathcal{D}^{(2)}((-1,1)) = \frac{1}{Z_2}\mathcal{D}^{(1)}((-1,1)) \cdot \sqrt{\frac{\varepsilon_1}{1-\varepsilon_1}} = \frac{1}{Z_2}\cdot\left(\frac{1+\varepsilon}{4}\right) \cdot \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \quad \searrow$$

$$\mathcal{D}^{(2)}((+1,-1)) = \frac{1}{Z_2}\cdot\left(\frac{1-\varepsilon}{4}\right) \cdot \sqrt{\frac{2+\varepsilon}{2-\varepsilon}} \quad \nearrow$$

$$\mathcal{D}^{(2)}((+1,+1)) = \frac{1}{Z_2}\cdot\frac{1}{4}\cdot\sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \quad \searrow$$

$$\mathcal{D}^{(2)}((-1,-1)) = \frac{1}{Z_2}\cdot\frac{1}{4}\cdot\sqrt{\frac{2+\varepsilon}{2-\varepsilon}} \quad \nearrow$$

We can find the value of Z_2 such that $\mathcal{D}^{(2)}$ is a probability distribution, meaning that sum of probability masses should be equal to 1.

$$D^{(2)}((-1,1)) + D^{(2)}((+1,-1)) + D^{(2)}((+1,+1)) + D^{(2)}((-1,-1)) = 1$$

$$\Rightarrow Z_2 = \{ \frac{1+\varepsilon}{4} \cdot \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} + \frac{1-\varepsilon}{4} \cdot \sqrt{\frac{2+\varepsilon}{2-\varepsilon}} + \frac{1}{4} \cdot \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} + \frac{1}{4} \cdot \sqrt{\frac{2+\varepsilon}{2-\varepsilon}}$$

$$= \frac{1}{4} \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \left((1+\varepsilon) + (1-\varepsilon) \cdot \left(\frac{2+\varepsilon}{2-\varepsilon}\right) + \frac{1}{\cancel{4}} + \frac{1}{\cancel{4}} \cdot \frac{2+\varepsilon}{2-\varepsilon} \right)$$

$$= \frac{1}{4} \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \cdot \frac{(1+\varepsilon)(2-\varepsilon) + (1-\varepsilon) \cdot (2+\varepsilon) + (2-\varepsilon) + 2+\varepsilon}{2-\varepsilon}$$

$$= \frac{1}{4} \cdot \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \cdot \frac{2+\varepsilon-\varepsilon^2+2-\varepsilon-\varepsilon^2+2-\varepsilon+2+\varepsilon}{2-\varepsilon}$$

$$= \frac{1}{4} \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \cdot \frac{8-2\varepsilon^2}{2-\varepsilon} = \frac{1}{4} \sqrt{\frac{2-\varepsilon}{2+\varepsilon}} \cdot \frac{2(2-\varepsilon)(2+\varepsilon)}{(2-\varepsilon)}$$

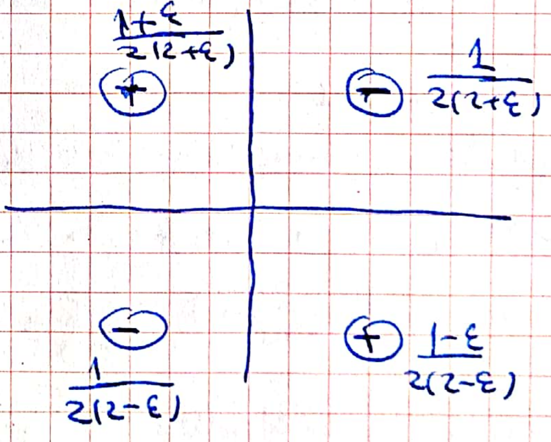
$$= \frac{1}{2} \cdot \sqrt{(2-\varepsilon)(2+\varepsilon)}$$

$$\Rightarrow D^{(2)}((-1,1)) = \frac{(1+\varepsilon)}{2(2+\varepsilon)}$$

$$D^{(2)}((+1,-1)) = \frac{(1-\varepsilon)}{2(2-\varepsilon)}$$

$$D^{(2)}((+1,+1)) = \frac{1}{2(2+\varepsilon)}$$

$$D^{(2)}((-1,-1)) = \frac{1}{2(2-\varepsilon)}$$



What is the error of the box classifier $h^3 = h_{1,0,+1}$ selected at round 1 on $D^{(2)}$? $Loss(h^3) = \frac{1}{2(2-\varepsilon)} + \frac{1-\varepsilon}{2(2-\varepsilon)} = \frac{2-\varepsilon}{2(2-\varepsilon)} = \frac{1}{2}$.

Round 2

- distribution $D^{(2)}$: $\begin{pmatrix} (-1,+1) & (1,-1) & (+1,+1) & (-1,-1) \\ \frac{1+\varepsilon}{2(2+\varepsilon)} & \frac{1-\varepsilon}{2(2-\varepsilon)} & \frac{1}{2(2+\varepsilon)} & \frac{1}{2(2-\varepsilon)} \end{pmatrix}$

- select the best classifier from $\mathcal{H}$, the one with the minimum empirical risk

$$L_{D^{(2)}}(h^1) = \frac{1}{2(2-\varepsilon)} + \frac{1}{2(2+\varepsilon)} = \frac{2+\varepsilon+2-\varepsilon}{2(2-\varepsilon)(2+\varepsilon)} = \frac{2}{(2-\varepsilon)(2+\varepsilon)} = \frac{4}{2(2+\varepsilon)(2-\varepsilon)}$$

$$L_{D^{(2)}}(h^2) = \frac{(1+\varepsilon)}{2(2+\varepsilon)} + \frac{1-\varepsilon}{2(2-\varepsilon)} = \frac{(1+\varepsilon)\times(2+\varepsilon)+(1-\varepsilon)(2+\varepsilon)}{2(2-\varepsilon)(2+\varepsilon)} = \frac{4-2\varepsilon^2}{2(2-\varepsilon)(2+\varepsilon)}$$

$$L_{D^{(2)}}(h^3) = \frac{1}{2} = \frac{4-\varepsilon^2}{2(2-\varepsilon)(2+\varepsilon)}$$

$$L_{D^{(2)}}(h^4) = \frac{1}{2} = \frac{4-\varepsilon^2}{2(2-\varepsilon)(2+\varepsilon)}$$

$$L_{D^{(2)}}(h^5) = \frac{(1+\varepsilon)}{2(2+\varepsilon)} + \frac{1}{2(2-\varepsilon)} = \frac{(1+\varepsilon)(2-\varepsilon)+2+\varepsilon}{2(2+\varepsilon)(2-\varepsilon)} = \frac{4+2\varepsilon-\varepsilon^2}{2(2+\varepsilon)(2-\varepsilon)}$$

$$L_{D^{(2)}}(h^6) = \frac{1}{2(2+\varepsilon)} + \frac{1-\varepsilon}{2(2-\varepsilon)} = \frac{(2-\varepsilon)+(1-\varepsilon)(2+\varepsilon)}{2(2+\varepsilon)(2-\varepsilon)} = \frac{4-2\varepsilon-\varepsilon^2}{2(2+\varepsilon)(2-\varepsilon)}$$

The smallest error is attained by $\underline{h^6}$. Thus it the weak-classifier selected at the current round.

So, for $t=2$ (round 2) we have $h_2 = h^6 = h_{2,0,-1}$

$$\varepsilon_2 = \frac{4-2\varepsilon-\varepsilon^2}{2(2+\varepsilon)(2-\varepsilon)}$$

$$w_2 = \frac{1}{2}\ln\left(\frac{1}{\varepsilon_2}-1\right) = \frac{1}{2}\ln\left(\frac{4-\varepsilon^2+2\varepsilon}{4-\varepsilon^2-2\varepsilon}\right)$$