COMP 585 **Spring 2024**

Probabilistic Toolkit for Learning and Computing *Instructor: Maryam Aliakbarpour*

4/4/24 *Scriber: Ria Stevens*

Lecture 26

---

1 Adaboost

- **Input** ϵ, δ, T
- Draw m samples, get training set $S = \{(x_1, y_1), ..., (x_m, y_m)\}$
- $D_1(i) = \frac{1}{m}$
- For $t = 1, ..., T$
  - $\hat{c}_t \leftarrow \text{WeakLearner}(\epsilon', \delta')$
  - update $D_{t+1}(i) \ \ \forall i$
- output H s.t. $H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t \hat{c}_t(x))$

1.1 Choosing $\epsilon_1, \epsilon_2, \delta', m, T$

- let $\mathcal{H}_T = \{H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t \hat{c}_t(x)) | \hat{c}_t \in \mathcal{C}\}$
- Recall $\hat{\text{err}}(H) \leq e^{-2\gamma^2 T} = \epsilon_1$
- Assume $\mathcal{H}_T$ has finite VC dimension.

  $\implies$ $\mathcal{H}_T$ satisfies uniform convergence for $m = O(\frac{1}{\epsilon_2^2} \text{VCdim}(\mathcal{H}) \log(2/\delta))$ samples

  $\implies$ $|\hat{\text{err}}(H) - \text{err}(H)| \leq \epsilon_2$

- We then have:

$$
\begin{aligned}
\text{err}(H) &= \text{err}(H) - \hat{\text{err}}(H) + \hat{\text{err}}(H) \\
&\leq |\text{err}(H) - \hat{\text{err}}(H)| + \hat{\text{err}} \\
&\leq \epsilon_2 + \epsilon_1
\end{aligned}
$$

- Choosing $\epsilon_1 = \epsilon_2 = \frac{\epsilon}{2}$, $\text{err}(H) \leq \epsilon$
- So, we have:

$$m = O(\frac{2}{\epsilon_2}\text{VCdim}(\mathcal{H})\log(2/\delta))$$
$$e^{-2\gamma^2 T} = \frac{\epsilon}{2} \implies T = O(\frac{\log(2/\epsilon)}{2\gamma^2})$$

- To choose δ', we have the probability that the weak learner succeeds T times is $T\delta'$. If we choose the probability of drawing enough samples from uniform convergence to be $\frac{\epsilon}{2}$, we want $T\delta' + \frac{\epsilon}{2} \leq \delta$, and can choose $\delta' = \frac{\delta}{2T}$.
- To complete our discussion of our choice of m, we need to determine the VC dimension of $\mathcal{H}$.

1.2 VC Dimension of $\mathcal{H}$

To analyze the VC dimension of $\mathcal{H}$, we first consider the VC Dimension of halfspaces.

We consider the set of halfspaces $\mathcal{G}$, where we define:

$$c_\alpha(x) = \text{sign}(\langle \alpha, x \rangle)$$
$$\mathcal{G} = \{c_\alpha | \alpha \in \mathbb{R}^d\ \}$$

Lemma 1.1. *VCdim($\mathcal{G}$) = d*

Proof. To prove Lemma 1.1, we need to show that $\mathcal{G}$ shatters a set of size d and that $\mathcal{G}$ cannot shatter any set of size $d + 1$.

We begin by proving that $\mathcal{G}$ shatters the set of unit vectors in d-dimensions. Let $S = \{\text{unit vectors } e_i\}_{i=1}^d$. Create a labelling of these vectors from $\mathcal{G}$. This gives us:

$$\langle \alpha, e_1 \rangle = y_1$$
$$\langle \alpha, e_2 \rangle = y_2$$
$$...$$
$$\langle \alpha, e_d \rangle = y_d$$

Let $\alpha = (y_1, y_2, ..., y_d)$. Then, c_α correctly labels all the unit vectors. So, $\mathcal{G}$ shatters the unit vectors.

Then, we show that no set of size $d + 1$ can be shattered by $\mathcal{G}$. Given $S = \{x_1, ..., x_m\}$, where $x_i \in \mathbb{R}^d$ and $m > d$, we have:

$$\textstyle\sum_{i=1}^m a_i x_i = \vec{0} \text{ for } a_i \in \mathbb{R}, \text{ where } \exists a_i \neq 0.$$

Let $I^+ = \{i | a_i > 0\}$ and $I^- = \{i | a_i \leq 0\}$.

Then, either $|I^+| \neq 0$ or $|I^-| \neq 0$. WLOG, assume $|I^-| \neq 0$.

Assume $\exists w \in \mathbb{R}^m$ that correctly labels each x_i.

$$\implies \langle w, x_i \rangle \geq 0 \ \forall i \in |I^+| \text{ and } \langle w, x_i \rangle < 0 \ \forall i \in |I^-|.$$

Then, we have:

$$\begin{aligned} 0 &\leq \sum_{i \in |I^+|} a_i \langle w, x_i \rangle \\ &= \langle w, \sum_{i \in |I^+|} a_i x_i \rangle \\ &= \langle w, \vec{0} - \sum_{i \in |I^-|} a_i x_i \rangle \\ &= \langle w, \sum_{i \in |I^-|} |a_i| x_i \rangle \\ &= \sum_{i \in |I^-|} |a_i| \langle w, x_i \rangle \\ &< 0 \\ &\text{Contradiction.} \end{aligned}$$

So, there does not exist w that correctly labels this set. □

Then, consider $\mathcal{G}'$, where α is defined only on the hypercube, $\{+1, -1\}^d$. That is, $\mathcal{G}' = \{c_\alpha | \alpha \in \{+1, -1\}^d\}$. Then, because $\mathcal{G}' \subset \mathcal{G}$, $\text{VCdim}(\mathcal{G}') \leq \text{VCdim}(\mathcal{G}) = d$.

Lemma 1.2. *VCdim*$(\mathcal{H}) \leq \theta(d_1 T)$, *where* $d_1 =$ *VCdim*(C).

Proof. We have m points $(x_1, ..., x_m)$, where $x_i \in \mathbb{R}^d$. We can map each point to the algorithm's output over time as:

$$x_i \xrightarrow{\hat{c} \in C} \{-1, 1\}^T \xrightarrow{\text{halfspace, } \hat{c}_\alpha \in \mathcal{G}} y_i = (\hat{c}_1(x_i), ..., \hat{c}_T(x_i))$$

This first jump has VCdim d_1 (that of C), and the second has VCdim T, as shown in the lemma. Intuitively, this will give us a VCdim $\in O(d_1 T)$

To show this more comprehensively, we use Sauer's Lemma. By Sauer's Lemma, $\forall S$ of size m, $|R_{\mathcal{G}}(S)| < (\frac{em}{d})^d$, where $d = \text{VCdim}(\mathcal{G})$. So, $|R_C(S)| \leq (\frac{em}{d_1})^{d_1}$.

If we fix $(\hat{c}_1, ..., \hat{c}_T)$, then the number of restrictions on $(\hat{c}_1(x_i), ..., \hat{c}_T(x_i))$ is less than $(\frac{em}{T})^T$.

Putting these together, we have that the number of restrictions on every possible combination of $\hat{c}_i$ is less than $(\frac{em}{T})^T (\frac{em}{d_1})^{d_1 T} \leq m^{(d+1)T}$.

So, for $m < \theta(dT \log(dT))$,

$$2^m \leq \# \text{ restrictions} \leq m^{(d+1)T}$$

□