Bias Variance trade off and PLA

Thursday, 19 January 2023 10:11 PM

VC Generalization Bound

$$E_{out}(g) \le E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4 m_H(2N)}{\delta}}$$

The VC bound is loose

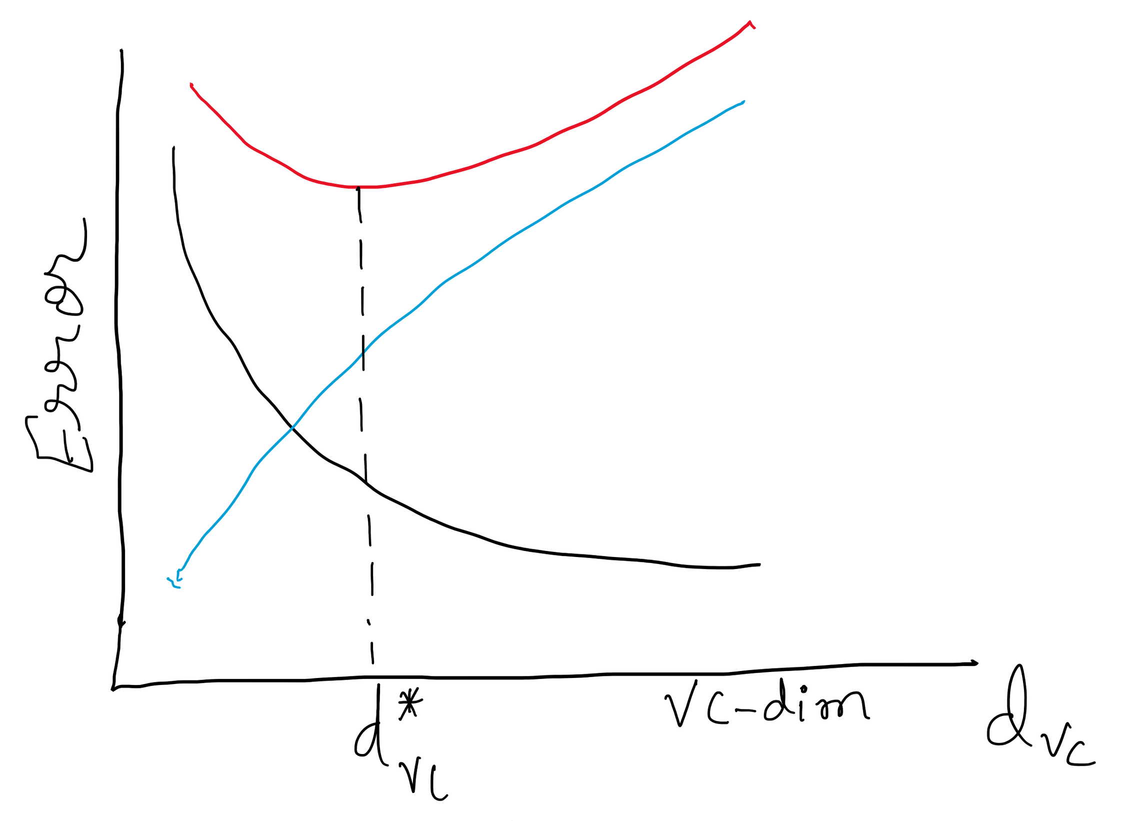
- Hoeffding's inequality has some slack
- $m_H(N)$ gives a worst case estimate
- Bounding $m_{\mathcal{H}}(N)$ by polynomial of order d_{vc}

Then what's the use?

- Establishes the feasability of learning for infinite hypothesis sets
- Loose but equally loose for all. Useful for comparing generalization performance of models.

Popular Rule of Thumb :-

$N \approx 10 \times d_{vc}$



Bias and Variance :-

$$E_{out}(g^{(D)}) = E_X\left[(g^{(D)}(x) - f(x))^2\right]$$

E_X denotes expected value w.r.t x

IMP :- $g^{(D)}$ is dependent on D

$$E_D\left[E_{out}(g^{(D)})\right]$$

$$= E_D\left[E_X\left[(g^{(D)}(x) - f(x))^2\right]\right]$$

$$= E_X\left[E_D\left[(g^{(D)}(x) - f(x))^2\right]\right]$$

$$= E_X\left[E_D\left[g^{(D)}(x)^2\right] - 2E_D\left[g^{(D)}(x)\right] f(x) + f(x)^2\right]$$

Let $E_D\left(g^{(D)}(x)\right) = \bar{g}(x)$

$$E_D\left[E_{out}(g^{(D)})\right]$$

$$= E_X\left[E_D\left[g^{(D)}(x)^2\right] - 2\bar{g}(x) f(x) + f(x)^2\right]$$

$$= E_X\left[\underbrace{E_D\left[g^{D}(x)^2\right] - \bar{g}(x)^2}_{E_D\left[(g^{(D)}(x) - \bar{g}(x))^2\right]} + \underbrace{\bar{g}(x)^2 - 2\bar{g}(x) f(x) + f(x)^2}_{(\bar{g}(x) - f(x))^2}\right]$$

$$\text{bias}(x) = \left(\bar{g}(x) - f(x)\right)^2$$

$$\text{Var}(x) = E_D\left[(g^{(D)}(x) - \bar{g}(x))^2\right]$$

$$\therefore E_D\left[E_{out}(g^{(D)})\right] = E_X\left[\text{bias}(x) + \text{Var}(x)\right]$$

→ Here along with $\mathcal{H}$ algorithm $\mathcal{A}$ also matters

---

The PLA Algorithm

$w^{(0)} = (0, 0, 0, 0, 0)$

While there is a misclassified point $x(t), y(t)$

$$w(t+1) = w(t) + y(t)x(t)$$