Perceptron's LB - Definitions

- Choose any linear prediction function

$$f^{\star}(\mathbf{x}) = \mathbf{u} \cdot \mathbf{x}$$

- The vector $\mathbf{u}$ can be chosen by examining

$$(\mathbf{x}_1, y_1)\,(\mathbf{x}_2, y_2)\ \ldots\ (\mathbf{x}_T, y_T)$$

- The hinge loss of $\mathbf{u}$ on $(\mathbf{x}_t, y_t)$

$$\ell_t^{\star} = \max\{0, 1 - y_t\,(\mathbf{u} \cdot \mathbf{x}_t)\}$$

- The 1-norm and 2-norm of the Hinge losses

$$D_1 = \sum_{t=1}^{T} \ell_t^{\star} \quad D_2 = \sqrt{\sum_{t=1}^{T} (\ell_t^{\star})^2}$$

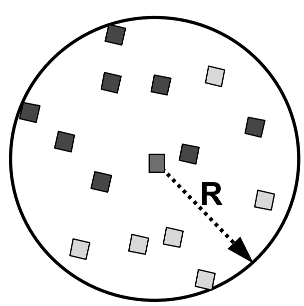
21

One More Assumption…

- All the input instances $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$

  are enclosed in a ball of radius R:

$$\forall t\ \|\mathbf{x}_t\| \leq R$$



22

Perceptron's Loss Bounds

$$M \leq \left(R\|\mathbf{u}\| + \sqrt{D_1}\right)^2$$

$$M \leq \left(R\|\mathbf{u}\| + D_2\right)^2$$

If the sample is separable ⇒ $M \leq R^2\|\mathbf{u}\|^2$

23

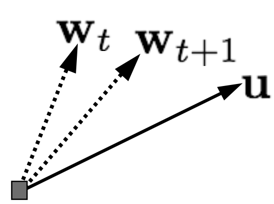
Proof - Intuition

- We will examine two terms:

$$\mathbf{w}_t \cdot \mathbf{u} \quad ; \quad \|\mathbf{w}_t\|$$

- Both terms grow as learning progresses but one grows faster…
- Overall, the following term increases monotonically as *t* grows

$$\frac{\mathbf{w}_t \cdot \mathbf{u}}{\|\mathbf{w}_t\|}$$



24

Proof

- ϵ_t an indicator denoting whether there was a mistake on round t ($\epsilon_t = 1$) or not ($\epsilon_t = 0$)

$$M = \sum_{t=1}^{T} \epsilon_t$$

$$\ell_t^\star = \max\{0, 1 - y_t \left(\mathbf{u} \cdot \mathbf{x}_t\right)\}$$

$$\ell_t^\star \geq 1 - y_t(\mathbf{u} \cdot \mathbf{x}_t) \;\Rightarrow\; y_t(\mathbf{u} \cdot \mathbf{x}_t) \geq 1 - \ell_t^\star$$

$$\begin{aligned} \mathbf{w}_{t+1} \cdot \mathbf{u} &= \mathbf{w}_t \cdot \mathbf{u} + y_t(\mathbf{x}_t \cdot \mathbf{u}) \\ &\geq \mathbf{w}_t \cdot \mathbf{u} + 1 - \ell_t^\star \end{aligned}$$

Proof (cont.)

- If there wasn't a prediction mistake:

$$\mathbf{w}_{t+1} \cdot \mathbf{u} = \mathbf{w}_t \cdot \mathbf{u}$$

- Summing up the two cases

$$\mathbf{w}_{t+1} \cdot \mathbf{u} = \epsilon_t - \epsilon_t \ell_t^\star$$

- Applying inequality+inequality multiple times:

$$\mathbf{w}_{t+1} \cdot \mathbf{u} \geq \sum_{s=1}^{t} \epsilon_s - \sum_{s=1}^{t} \epsilon_s \ell_s^\star = M - \sum_{s=1}^{t} \epsilon_s \ell_s^\star$$

Proof (cont.)

- Norm of $\mathbf{w}_t$ grows in case of a prediction mistake

$$\begin{aligned}\|\mathbf{w}_{t+1}\|^2 &= \mathbf{w}_{t+1} \cdot \mathbf{w}_{t+1} \\ &= (\mathbf{w}_t + y_t \mathbf{x}_t) \cdot (\mathbf{w}_t + y_t \mathbf{x}_t) \\ &= \|\mathbf{w}_t\|^2 + \|\mathbf{x}_t\|^2 + 2y_t(\mathbf{w}_t \cdot \mathbf{x}_t)\end{aligned}$$

$$\leq R \qquad \leq 0$$

- Prediction mistake ➔ $\|\mathbf{w}_{t+1}\|^2 \leq \|\mathbf{w}_t\|^2 + R$

- Summing over all rounds from *s* through *t*

$$\|\mathbf{w}_{t+1}\|^2 \leq MR^2 \;\Rightarrow\; \|\mathbf{w}_{t+1}\| \leq \sqrt{M}R$$

27

Proof (cont.)

- From Cauchy-Schwartz

$$\mathbf{w}_{t+1} \cdot \mathbf{u} \;\leq\; \|\mathbf{w}_{t+1}\| \; \|\mathbf{u}\|$$

$$M - \sum_{s=1}^{t} M - \sum_{s=1}^{t} \epsilon_s \ell_s^\star \leq \sqrt{M}\, R\, \|\mathbf{u}\| \; \overline{M}R\|\mathbf{u}\|$$

CS Ineq.

$$\leq \sqrt{\sum_{s=1}^{t} \epsilon_s^2} \sqrt{\sum_{s=1}^{t} (\ell_s^\star)^2} \;=\; \sqrt{M} D_2$$

28

Prooof

- Finally…

$$M - \sqrt{M} D_2 \leq \sqrt{M} R \|\mathbf{u}\|$$

- Divide by $\sqrt{M}$ and move terms around

$$\sqrt{M} \leq R\|\mathbf{u}\| + D_2$$

$$\Downarrow$$

$$M \leq (R\|\mathbf{u}\| + D_2)^2 \quad \blacksquare$$

- For D_1 use Holder’s inequality instead of CS

29