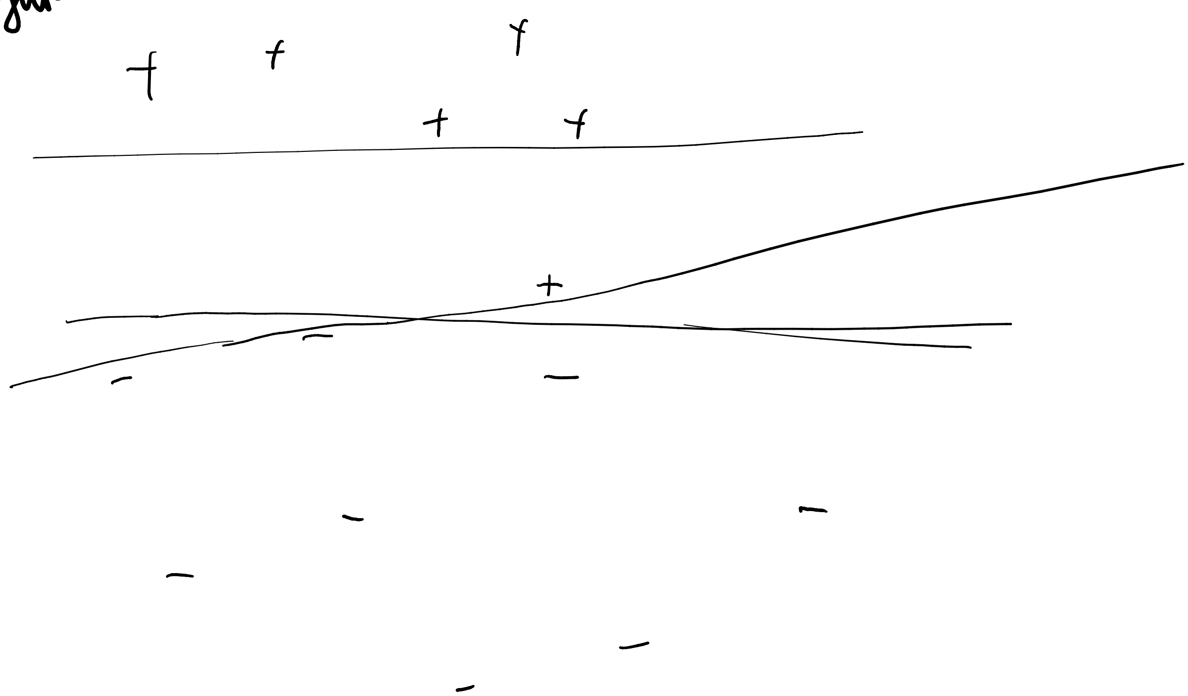
Margins and SVM

Wednesday, October 27, 2021 7:14 AM

YuRan



$$w \cdot x_i \geq 1 - \varepsilon_i \qquad \ell(x_i) = +$$

$$w \cdot x_i \leq -1 + \varepsilon_i \qquad \ell(x_i) = -$$

$$\varepsilon_i \geq 0$$

$$\text{Min} \quad \sum \varepsilon_i \quad \longleftarrow \text{"Hinge Loss"}$$

OPT $=0 \Rightarrow$ perfect separator. preferred by this program.

$$\gamma = \min_{x} \frac{|w^* \cdot x|}{\|x\|}$$

$$\gamma = \min_x \frac{|w^* \cdot x|}{\|w^*\|}$$

$$\|w^*\| = \frac{\min |w^* \cdot x|}{\gamma} = \frac{1}{\gamma} \quad \text{above.}$$

So how about

$$\text{Min} \quad \|w\|^2 + C \sum_i \varepsilon_i$$

$$w \cdot x_i \geq 1 - \varepsilon_i \qquad \ell(x_i) = +$$

$$w \cdot x_i \leq -1 + \varepsilon_i \qquad \ell(x_i) = -$$

rewards large margin. Support Vector Machine.
How to choose C? depends on data / application.

Perceptron gives such a guarantee!

Thm. # mistakes of Perceptron $= \min_w \left(\frac{1}{\gamma^2} + 2 \,(\text{Hinge Loss}) \right)$

Pf. Consider $\frac{w \cdot w^*}{\|w\|}$

For each mistake $w \leftarrow w + \ell(x) x_i$

For each mistake $w \leftarrow w + \ell(x) x_i$

$$w \cdot w^* \leftarrow w \cdot w^* + \ell(x_i)(w^* \cdot x_i)$$

increases by at least $1 - \varepsilon_i$

So after M mistakes total increase $\geq M - \sum \varepsilon_i$

$$\geq M - L$$

L = total hinge loss.

Meanwhile $\|w\|^2 \leftarrow \|w\|^2 + \|x_i\|^2 + 2\ell(x)(w \cdot x_i)$

$$\leq \|w\|^2 + 1 \quad + \quad \leq 0.$$

So $\|w\|^2 \leq M$.

$$|w \cdot w^*| \leq \|w\| \, \|w^*\|$$

$$\Rightarrow \quad M - L \leq \|w^*\| \sqrt{M}$$

$$M^2 + L^2 - 2ML \leq \frac{M}{\gamma^2}$$

$$M \leq \frac{1}{\gamma^2} + 2L - \frac{L^2}{M}$$

$$\boxed{M \leq \frac{1}{\gamma^2} + 2L}$$

$$\gamma^2$$

Random Projection and Margins

$$R = \frac{1}{k} \cdot \begin{array}{cc} & k \\ n & \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix} \end{array} \qquad R_{ij} \sim N(0,1)$$

$$\mathbb{E}\, \|R^T x\|^2 = \|x\|^2$$

<u>Lemma</u>.

$$\Pr\left(\left| \|R^T x\|^2 - \|x\|^2 \right| > \varepsilon \|x\|^2 \right) \leq 2 e^{-\left(\frac{\varepsilon^2 - \varepsilon^3}{4}\right) \cdot k}$$

$$x' = R^T x \quad y' = R^T y$$

<u>Lemma</u>.

$$\Pr\left(|x' \cdot y' - x \cdot y| > \varepsilon \|x\| \|y\| \right) \leq 2 e^{-c \varepsilon^2 k}.$$

RP preserves margin approximately.

$$\gamma \longrightarrow \geq \frac{\gamma}{2} \quad \text{WHP}.$$

In $\mathbb{R}^k$, need only $m = O\left(\frac{k}{\varepsilon} \left(\log \frac{1}{\varepsilon} + \log \frac{1}{\delta} \right) \right)$

samples. $k = O\left(\frac{1}{\gamma^2} \log\left(\frac{1}{\gamma \varepsilon \delta} \right) \right)$ suffices to

$\gamma^2 \quad (\gamma \varepsilon \delta))$

preserve $\frac{\gamma}{2}$ margin on m samples.

Now run Perceptron (or any linear separator algo) in $\mathbb{R}^k$. Faster, more efficient!