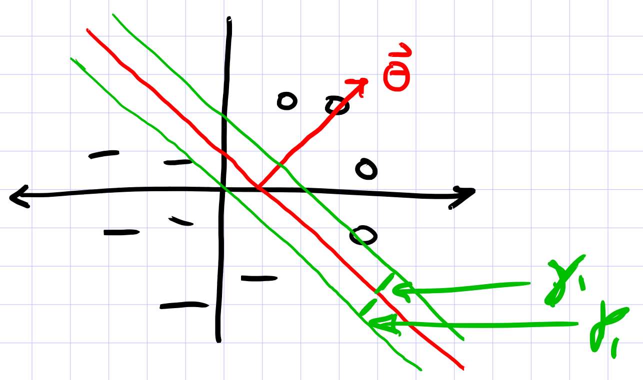
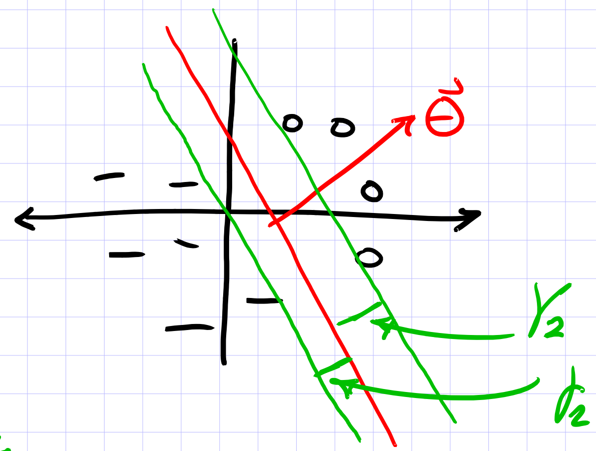
Lecture 11 : 2/22/17





$\gamma_1 < \gamma_2$

Perceptron Mistake Bound

$$\vec{\theta}^{k+1} = y^{(1)}\vec{x}^{(1)} + y^{(2)}\vec{x}^{(2)} + \cdots + y^{(k)}\vec{x}^{(k)}$$

$$\|\theta^{k+1}\| = \|\theta^{k+1}\|\|\theta^*\| \geq \vec{\theta}^{k+1}\cdot\vec{\theta}^* = y^{(1)}x^{(1)}\cdot\theta^* + \cdots + y^{(k)}\vec{x}^{(k)}\cdot\theta^*$$

$$= \gamma + \cdots + \gamma$$

$$= k\gamma$$

$$\|w\|\cdot\|u\| \geq w\cdot u$$

Part I

$$\|\theta^{k+1}\|^2 = \|y^{(1)}x^{(1)} + \cdots + y^{(k)}x^{(k)}\|^2$$

$$\leq \|x^{(1)}\|^2 + \cdots + \|x^{(k)}\|^2$$

$$\leq R^2 + \cdots + R^2$$

$$\leq kR^2$$

$$\Rightarrow \|\theta^{k+1}\| \leq R\sqrt{k}$$

Part II

Kernel Perceptron

Standard Perceptron Alg. (Online)

For $t = 1, 2, \ldots$

- Receive $(\vec{x}^{(t)}, y^{(t)})$
- Predict $\hat{y} = \text{sign}(\vec{\Theta}^T \vec{x}^{(t)})$
- If $\hat{y} \neq y^{(t)}$:
  update $\vec{\Theta} \leftarrow \vec{\Theta} + y^{(t)} \vec{x}^{(t)}$

Notice: After t examples were received

$$\vec{\Theta} = \alpha_1 x^{(1)} + \alpha_2 x^{(2)} + \cdots + \alpha_t x^{(t)}$$

where $\alpha_i = \begin{cases} y^{(i)} & \text{if mistake} \\ 0 & \text{otherwise} \end{cases}$

$$\begin{aligned}
\vec{\Theta}^T \vec{x} &= (\alpha_1 x^{(1)} + \cdots + \alpha_t x^{(t)})^T \vec{x} \\
&= \alpha_1 (x^{(1)})^T \vec{x} + \cdots + \alpha_t (x^{(t)})^T \vec{x} \\
&= \sum_{i=1}^{t} \alpha_i (x^{(i)})^T \vec{x} \\
&= \sum_{i=1}^{t} \alpha_i \, k(\vec{x}^{(i)}, \vec{x})
\end{aligned}$$

this is called the "kernel trick"

where $k(x, z) = x^T z$

Kernel Perceptron Alg.

For $t = 1, 2, \ldots$

- Receive $(\vec{x}^{(t)}, y^{(t)})$
- Predict
  $\hat{y} = \text{sign}\left(\sum_{i=1}^{t-1} \alpha_i \, k(\vec{x}^{(i)}, \vec{x}^{(t)})\right)$
- If $\hat{y} \neq y^{(t)}$
  $\alpha_t = y^{(t)}$
  Else:
  $\alpha_t = 0$

☆ Computational trick:
only store examples $x^{(i)}$ for which $\alpha_i \neq 0$

Closure Properties

k_1 and k_2 are known kernels

$$k(x,z) \triangleq k_1(x,z) + k_2(x,z)$$

$$k(x,z) \triangleq k_1(x,z)\, k_2(x,z)$$

$$k(x,z) \triangleq c \cdot k_1(x,z) \qquad \forall c > 0$$

$$k(x,z) \triangleq k_1(x,z) + c$$

$$k(x,z) \triangleq \exp(k_1(x,z))$$

$$\vdots$$

k is a kernel iff

$$\exists\, \phi : \mathbb{R}^M \to \mathbb{R}^D \quad \text{s.t.} \quad k(x,z) = \phi(x)^T \phi(z)$$