Labelled points with $x_1, \ldots, x_n$.

Hyperplane separator.

Margins.

Inside unit ball.

Margin $\gamma$.

Hyperplane: $w \cdot x \geq \gamma$ for + points.

$w \cdot x = \langle w, x \rangle$ for - points.

Put points on unit ball.

$w \cdot x = \cos \theta$ Will assume positive labels!

negate the negative.

**Claim 1:** $w_{t+1} \cdot w \geq w_t \cdot w + \gamma$.

**Claim 2:** $\|w_{t+1}\|^2 \leq \|w_t\|^2 + 1$

$M$-number of mistakes in algorithm.

$\gamma M \leq w_{t+1} \cdot w \leq \|w_t\| \leq \sqrt{M}$.

$M \leq \frac{1}{\gamma^2}$

---

**Perceptron Algorithm**

An aside: a hyperplane is a perceptron.

(single layer neural network.)

Alg: Given $x_1, \ldots, x_n$.

Let $w_1 = x_1$.

For each $x_i$, $w_t \cdot x_i$ is wrong sign (negative)

$w_{t+1} = w_t + x_i$

$t = t + 1$

**Theorem:** Algorithm only makes $1/\gamma^2$ mistakes.

Idea: Mistake on positive $x_i$:

$w_{t+1} \cdot x_i = (w_t + x_i) \cdot x_i = w_t x_i + 1$.

A step in the right direction!

**Claim 1:** $w_{t+1} \cdot w \geq w_t \cdot w + \gamma$.

A $\gamma$ in the right direction!

Mistake on positive $x_i$;

$w_{t+1} \cdot w = (w_t + x_i) \cdot w = w_t \cdot w + x_i \cdot w$

$\geq w_t \cdot w + \gamma$.

---

**Hinge Loss.**

Most of data has good separator.

**Claim 1:** $w_{t+1} \cdot w \geq w_t \cdot w + \gamma$.

Don't make progress or tilt the wrong way.

How much bad tilting?

Rotate points to have $\gamma$-margin.

Total rotation: $TD$.

Analysis: subtract bad tilting part.

**Claim 1:** $w_{t+1} \cdot w \geq w_t \cdot w + \gamma$ rotation for $x_i$.

$w_{t+1} \cdot w \geq w_t \cdot w + \gamma$.

$M \cdot 2TD_{y} \leq \sqrt{M}$

Quadratic equation: $\gamma^2M^2 - (2\gamma TD_{y} + 1)M + TD_{y}^2 \leq 0$.

Uh...

One implication: $M \leq \frac{1}{\gamma^2} + \frac{1}{\gamma TD_{y}}$.

The extra is (twice) the amount of rotation in units of $\gamma$.

Hinge loss: $\frac{1}{\gamma^2} TD_{y}$.
Approximately Maximizing Margin Algorithm

There is a $\gamma$ separating hyperplane.
Find it! (Kind of.)
Any point within $\gamma/2$ is still a mistake.
Let $w_1 = x_1$.
For each $x_2, \ldots, x_n$, if $w_1 \cdot x_i < \gamma/2$, $w_{t+1} = w_t + x_i$, $t = t + 1$
Claim 1: $w_{t+1} \cdot w \geq w_t \cdot w + \gamma$
2.
Same (ish) as before.

Kernel Functions.

Map $x$ to $\phi(x)$.
Good separator for points under $\phi(.)$. Problem: complexity of computing in higher dimension.
Recall perceptron. Only compute dot products!
Test: $w_t \cdot x_i > \gamma$
$w_t = x_1 + x_2 + x_3 \cdots$
Support Vectors: $x_1, x_2, \ldots$
$\rightarrow$ Support Vector Machine.
Kernel trick: compute dot products in original space.
Kernel function for mapping $\phi(.)$: $K(x, y) = \phi(x) \cdot \phi(y)$
$K(x, y) = (1 + x \cdot y)^d \phi(x) = [1, \ldots, x_i, \ldots, x_j \cdots].$ Polynomial.
$K(x, y) = (1 + x_1 y_1)(1 + x_2 y_2) \cdots (1 + x_n y_n)$
$\phi(x)$ - product of all subsets.
$K(x, y) = \exp(C|x - y|^2)$ Infinite dimensional space. Gaussian Kernel.

Margin Approximation: Claim 2

Claim 2(?) $|w_{t+1}|^2 \leq |w_t|^2 + 1$?
Adding $x_i$ to $w_i$ even if in correct direction.
Obtuse triangle.
$|v|^2 \leq |w|^2 + \frac{1}{2} \gamma^2$ (square right hand side.)
Red bit is at most $\gamma^2/2$.
Together: $|w_{t+1}| \leq |w_t| + \frac{\gamma}{\sqrt{2}} + \frac{1}{2}$

Other fat separators?

No hyperplane separator. Circle separator!
Map points to three dimensions.
map point $(x, y)$ to point $(x, y, x^2 + y^2)$.
Hyperplane separator in three dimensions.

Support Vector Machine

Pick Kernel.
Run algorithm that:
(1) Uses dot products.
(2) Outputs hyperplane that is linear combination of points.
Perceptron.
Max Margin Problem as Convex optimization:
$\min |w|^2$ where $\forall i \ w \cdot x_i \geq 1$.

Video

“http://www.youtube.com/watch?v=3liCbRZPrZA”
See you on Tuesday.