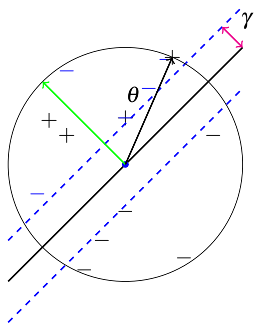
Today

Perceptron.

Support Vector Machine.



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

Hyperplane separator.

Margins.

Inside unit ball.
Margin γ
Hyperplane:
  $w \cdot x \geq \gamma$ for $+$ points.
  $w \cdot x \leq -\gamma$ for $-$ points.
Put points on unit ball.
  $w \cdot x = cos\theta$ Will assume
positive labels!
  negate the negative.

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 $\frac{1}{\gamma}^2$ 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 γ in the right direction!

Mistake on positive x_i;
  $w_{t+1} \cdot w = (w_t + x_i) \cdot w = w_t \cdot w + x_i \cdot w$
  $\geq w_t \cdot w + \gamma$.

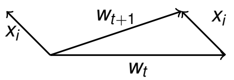
□

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

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



$w_{t+1} = w_t + x_i$
Less than a right angle!
$\rightarrow |w_{t+1}|^2 \leq |w_t|^2 + |x_i|^2 \leq |w_t|^2 + 1$.
Algebraically.
  Positive x_i, $w_t x_i \leq 0$.
  $(w_t + x_i)^2 = |w_t|^2 + 2w_t \cdot x_i + |x_i|^2$.
  $\leq |w_t|^2 + |x_i|^2 = |w_t|^2 + 1$.

□

Claim 2 holds even if no separating hyperplane!

Putting it together...

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.

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

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

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 γ-margin.
  Total rotation: TD_γ.
    Anaylsis: subtract bad tilting part.

Claim 1: $w_{t+1} \cdot w \geq w_t \cdot w + \gamma -$ rotation for x_{i_t}.

$w_M \geq \gamma M - TD_\gamma$ + Claim 2. $\rightarrow \gamma M - TD_\gamma \leq \sqrt{M}$

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

Uh...

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

The extra is (twice) the amount of rotation in units of γ.

Hinge loss: $\frac{1}{\gamma} TD_\gamma$.

Approximately Maximizing Margin Algorithm

There is a γ separating hyperplane.

Find it! (Kind of.)

Any point within $\gamma/2$ is still a mistake.

Let $w_1 = x_1$,

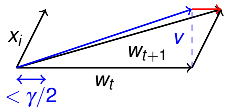
For each $x_2, \ldots x_n$,
if $w_t \cdot x_i < \gamma/2$, $w_{t+1} = w_t + x_i$, $t = t+1$

Claim 1: $w_{t+1} \cdot w \geq w_t w + \frac{\gamma}{2}$.

Same (ish) as before.

Margin Approximation: Claim 2

Claim 2(?): $|w_{t+1}|^2 \leq |w_t|^2 + 1$??



Adding x_i to w_t even if in correct direction.

Obtuse triangle.

$|v|^2 \leq |w_t|^2 + 1$
$\rightarrow |v| \leq |w_t| + \frac{1}{2|w_t|}$
(square right hand side.)

Red bit is at most $\gamma/2$.

Together: $|w_{t+1}| \leq |w_t| + \frac{1}{2|w_t|} + \frac{\gamma}{2}$

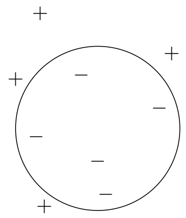
If $|w_t| \geq \frac{2}{\gamma}$, then $|w_{t+1}| \leq |w_t| + \frac{3}{4}\gamma$.

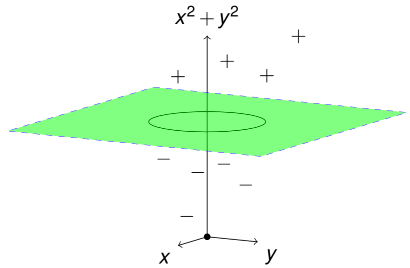
M updates $|w_M| \leq \frac{2}{\gamma} + \frac{3}{4}\gamma M$.

Claim 1: Implies $|w_M| \geq \gamma M/2$.

$\gamma M/2 \leq \frac{2}{\gamma} + \frac{3}{4}\gamma M \rightarrow M \leq \frac{8}{\gamma^2}$

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.

Kernel Functions.

Map x to $\phi(x)$.

Good separator for points under $\phi(\cdot)$.

Problem: complexity of computing in higher dimension.

Recall perceptron. Only compute dot products!
Test: $w_t \cdot x_i > \gamma$
$w_t = x_{i_1} + x_{i_2} + x_{i_3} \cdots$

Support Vectors: $x_{i_1}, x_{i_2}, \ldots$
$\rightarrow$ Support Vector Machine.

Kernel trick: compute dot products in original space.

Kernel function for mapping $\phi(\cdot)$: $K(x,y) = \phi(x) \cdot \phi(y)$

$K(x,y) = (1 + x \cdot y)^d$ $\phi(x) = [1, \ldots, x_i, \ldots, x_i x_j \ldots]$. 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.

Video

"http://www.youtube.com/watch?v=3liCbRZPrZA"

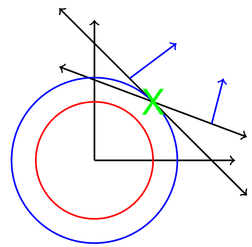
Support Vector Machine

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 $\forall i$ $w \cdot x_i \geq 1$.



Algorithms output: tight hyperplanes!

See you on Tuesday.