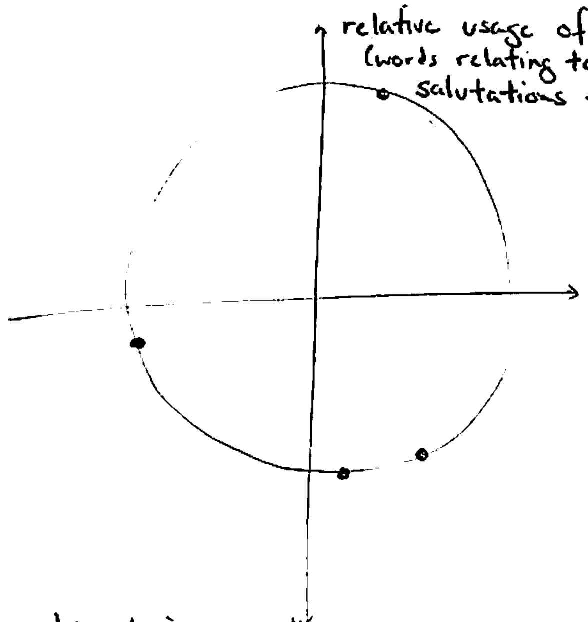
Quick tutorial: geometric intuitions on (text) classification

Recall the tf-idf vector representation for docs from last time.

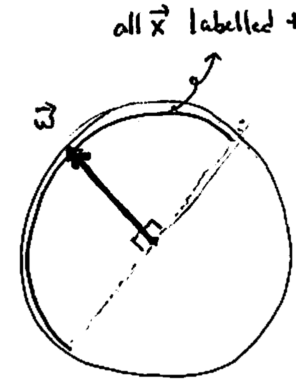
$\vec{x}_i$ = rep of i^{th} doc = $(\vec{x}_i^{(1)}, x_i^{(2)}, \ldots)$

For visualization purposes, it's useful to assume we subtract off the corpus-wide average of each coordinate from each coordinate, allowing negative values.

Example for ~~text~~ email spam classification:



one set of possible decision rules:
~~the class of lin separating~~
"dividing lines passing through the origin", (plus direction for positive class) which are linear separators, representable by ~~their~~ a unit normal vector $\vec{w}$.



The decision rule is then given by:

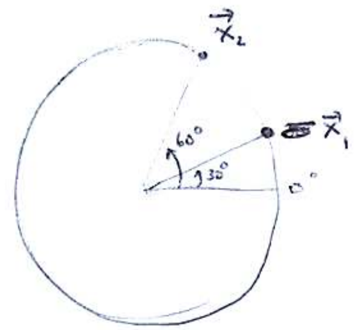
~~class = sign(~~

$$\text{class}(\vec{x}_i) = \text{sign}(\vec{w}\cdot\vec{x}_i)$$

> recall $\vec{w}\cdot\vec{x}$: $\|\vec{w}\|\|\vec{x}\|\cos(\angle(\vec{w},\vec{x}))$
>
> In our case, reduces to $\cos(\angle(\vec{w},\vec{x}_i))$
> ≥ 0 iff $\vec{x}_i$ is "within" 90° either way of $\vec{w}$, in 2-d.

(note: typical SVM formulation converts to non-unit-length $\vec{w}'$, ~~decision rule sign~~ and threshold $b \in \mathbb{R}$, decision rule $\text{sign}(\vec{w}'\cdot x + b)$.)

"Quiz": consider this situation, with two as-yet unlabeled points (documents)



(a) What is the set of such classifiers that would give both $\vec{x}_1$ & $\vec{x}_2$ positive labels?

(b) What are the sets of equivalence classes (each class a set of classifiers) where equivalence = assigns same labels to the points?
(You figured out one such class in (a)).

So: in some sense, there are only four (equivalence classes of) classifier to pick from, not infinitely many.
In general, ≤ # of possible labelings (not all may be achievable), each equivalence class of classifiers = distinct region on hypersphere.
(see Fig 8.1 of Vapnik '98)