Suppose we have sample points $X_1, X_2, \ldots, X_N$ which are represented as row vectors. For each sample point, the label $y_i = \{1 \quad if, X_i \in classC, and \ -1 \quad if, X_i \notin classC.$ Goal: find the weights ω such that $X_i \cdot \omega \geq 0 \qquad if \qquad y_i = 1 \ \ X_i \cdot \omega \leq 0 \qquad if \qquad y_i = -1.$ [$X_i \cdot \omega$ is the inner product which is the signed distance] Equivalently: $y_i X_i \cdot \omega \geq 0.$ Define the *loss function* $L(z, y_i) = \{0 \quad if \quad y_i z \geq 0, and \ -y_i z \quad otherwise.$ Define *rise function* (aka *objective function* or *cost function*) $R(\omega) = \frac{1}{n}\sum_{i=1}^{n} L(X_i \cdot \omega, y_i) = \frac{1}{n}\sum_{i \in V} -y_i X_i \cdot \omega$ where$V$is the set of indices i for which $y_i X_i \cdot \omega < 0$, essentially V contains the indices of samples which are incorrectly classified.

Point x lies on hyperplane {$z: \omega \cdot z = 0$} $\leftrightarrows \omega \cdot x = 0 \leftrightarrows \omega$ lies on hyperplane {$z: x \cdot z = 0$} **Draw a pic to illustrate this** For a sample point x in class C, ω and x must be on same side of the hyperplane that x transforms into. For a point not in class C, ω and x must be on opposite sides of the hyperplane that x transforms into. **Draw a pic to illustrate this** We get an optimization problem; We need a optimization algorithm to solve it.

Gradient descent

balabalabala ==I will fill in this part later.==

$\omega \leftarrow$ arbitrary nonzero starting point while $R(\omega) > 0$

- $V \leftarrow$ set of indices i for which $y_i X_i \cdot \omega < 0$
- $\omega \leftarrow \omega + \epsilon \sum_{i \in V} y_i X_i$

return ω ϵ is step size aka learning rate. Problem: Slow! We can use Stochastic gradient descent(only choose on point from V)

Maximum Margin Classifiers

The margin of a linear classifier is the distance from the decision boundary to the nearest sample point. Our goal is to make the margin as wide as possible. We enforce the **constraints**

- $y_i(\omega \cdot X_i + \alpha) \geq 1$

I don't know why the right-hand side is 1, but I think since we have α it can be set to any value

$||\omega|| = 1$, signed distance from hyperplane to X_i is $\omega \cdot X_i + \alpha$. Otherwise, it's$\frac{\omega}{||\omega||} \cdot X_i + \frac{\alpha}{||\omega||}$. **There are two ways to illustrate why this is signed distance, I will draw two pic to illustrate this later.** Hence the margin is $\min_i \frac{1}{||\omega||}|\omega \cdot X_i + \alpha| \geq \frac{1}{||\omega||}$, consider the constraints we can easily get this inequality. To maximize the margin $\rightarrow$ maximize $||\omega||$ $\rightarrow$ minimize$||\omega||$. We have optimization problem:

Find ω and α that minimize $||\omega||^2$ subject to $y_i(X_i \cdot \omega) \geq 1$ for all $i \in [1, n]$