

Cupcakes versus muffins

Support vector machines

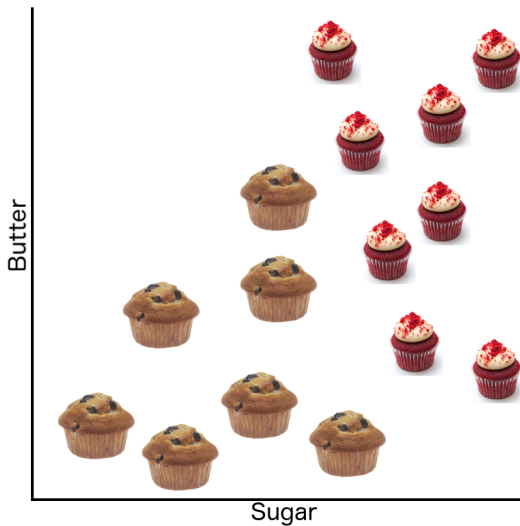
Arthur J. Parzygnat

University of Connecticut
SIGMA

January 26, 2018

- 1 Cupcakes and muffins
 - What's the difference?
- 2 Hyperplanes
 - Parametrizing hyperplanes
 - Marginal hyperplanes
 - The definition of a support vector machine
- 3 Existence and uniqueness of SVMs
 - Enlarging margins
 - Existence of SVMs
- 4 A simple example of an SVM

What's the difference?



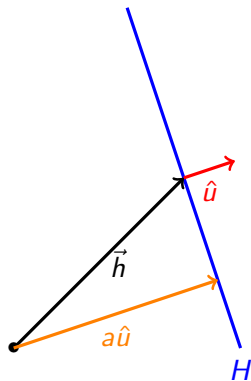
Parametrizing hyperplanes

Lemma 1

Let $H \subseteq \mathbb{R}^n$ be a hyperplane. Then there exists a vector $\vec{w} \in \mathbb{R}^n \setminus \{\vec{0}\}$ and a real number $c \in \mathbb{R}$ such that H is the solution set of $\langle \vec{w}, \vec{x} \rangle - c = 0$.

Proof.

Let $\vec{h} \in H$. Then $H - \vec{h}$ is an $(n-1)$ -dimensional subspace in \mathbb{R}^n . Hence, $(H - \vec{h})^\perp$ is a one-dimensional subspace spanned by some normalized vector \hat{u} . Because $\text{span}\{\hat{u}\}$ is perpendicular to H , there exists an $a \in \mathbb{R}$ such that $a\hat{u} \in H$. Then, H is the solution set of $\langle \hat{u}, \vec{x} \rangle - a = 0$. \square



Marginal planes

Note that a hat over a vector signifies that it is a unit vector (has magnitude 1). There are many ways to describe hyperplanes. It should not be obvious why it is preferable to use the description given by a vector and a number. Furthermore, notice that there are redundancies with this parametrization. For example, $(\lambda \vec{w}, \lambda c)$ and (\vec{w}, c) describe the same hyperplane for all $\lambda \in \mathbb{R} \setminus \{0\}$.

Definition 2

Let $\vec{w} \in \mathbb{R}^n \setminus \{\vec{0}\}$ and $c \in \mathbb{R}$ with associated plane H given by the solution set of $\langle \vec{w}, \vec{x} \rangle - c = 0$. The marginal planes H_+ and H_- associated to (\vec{w}, c) are the solution sets to $\langle \vec{w}, \vec{x} \rangle - c = 1$ and $\langle \vec{w}, \vec{x} \rangle - c = -1$, respectively, i.e.

$$H_{\pm} := \{ \vec{x} \in \mathbb{R}^n : \langle \vec{w}, \vec{x} \rangle - c = \pm 1 \}.$$

Marginal planes example

Example 3

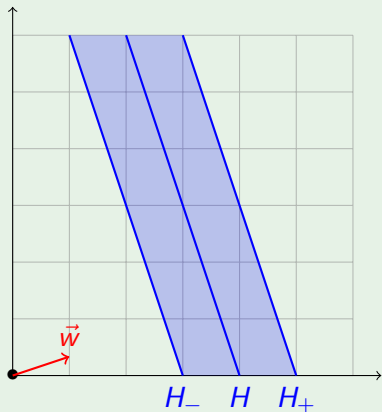
In \mathbb{R}^2 , if $\vec{w} = \frac{1}{3} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $c = 4$, then the marginal planes would look like the figure on the right since they are described by

$$H : \frac{1}{3} (3x + y) - 4 = 0$$

$$H_+ : \frac{1}{3} (3x + y) - 4 = 1$$

$$H_- : \frac{1}{3} (3x + y) - 4 = -1$$

Notice that the vector $\left(\frac{c}{\|\vec{w}\|^2} \right) \vec{w} = \frac{6}{5} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ lies on the plane H .



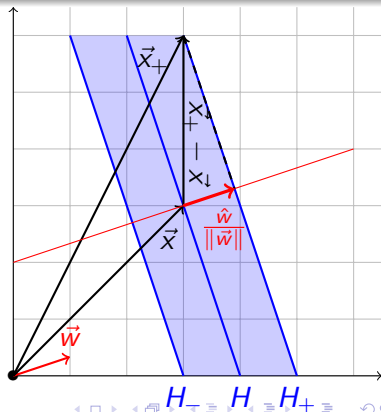
The distance between marginal hyperplanes

Lemma 4

Let $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ describe a hyperplane H . The perpendicular distance between H and H_{\pm} is $\frac{1}{\|\vec{w}\|}$.

Proof.

Let $\vec{x}_+ \in H_+$ and $\vec{x} \in H$. The orthogonal distance between H and H_+ is given by $\langle \vec{x}_+ - \vec{x}, \frac{\vec{w}}{\|\vec{w}\|} \rangle = \frac{1}{\|\vec{w}\|} (\langle \vec{w}, \vec{x}_+ \rangle - \langle \vec{w}, \vec{x} \rangle) = \frac{1}{\|\vec{w}\|} ((1+c) - c) = \frac{1}{\|\vec{w}\|}$ by the definition of H and H_+ in terms of (\vec{w}, c) . A similar calculation holds for H_- . \square



The margin and margin width

Definition 5

Let $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ describe a hyperplane H . The convex region between H_+ and H_- is called the margin of (\vec{w}, c) . The orthogonal distance between H_+ and H_- , which is given by $\frac{2}{\|\vec{w}\|}$, is called the margin width of (\vec{w}, c) .

Even though (\vec{w}, c) can be scaled to $(\lambda\vec{w}, \lambda c)$ to give the same H for any $\lambda \in \mathbb{R} \setminus \{0\}$, notice that the marginal planes are different. This is because the margin width has scaled by a factor of $\frac{1}{\lambda}$.

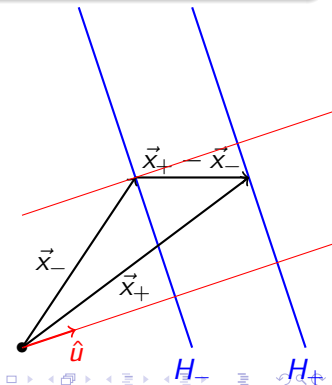
Parametrizing hyperplanes and their margins

Lemma 6

An ordered pair of distinct parallel hyperplanes (H_-, H_+) in \mathbb{R}^n determine a unique $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ whose marginal planes agree with H_- and H_+ .

Proof.

Let $\vec{x}_+ \in H_+$ and pick $\hat{u} \in (H_+ - \vec{x}_+)^\perp$ such that if $\lambda \hat{u} \in H_-$ and $\mu \hat{u} \in H_+$, then $\lambda < \mu$ (i.e. \hat{u} is perpendicular to H_+ and points from H_- to H_+). Also, let $\vec{x}_- \in H_-$. Since the orthogonal separation between the planes H_+ and H_- is $\langle \vec{x}_+ - \vec{x}_-, \hat{u} \rangle$, set $\vec{w} := \left(\frac{2}{\langle \vec{x}_+ - \vec{x}_-, \hat{u} \rangle} \right) \hat{u}$ and $c := \frac{\langle \vec{u}, \vec{x}_+ + \vec{x}_- \rangle}{\langle \vec{u}, \vec{x}_+ - \vec{x}_- \rangle}$. Then (\vec{w}, c) has H_+ & H_- as its marginal planes. \square



Parametrizing hyperplanes and their margins

Corollary 7

There is a 1-1 correspondence between the set of (ordered) pairs of parallel hyperplanes (the marginal hyperplanes) and the set $(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$.

Henceforth, we will abuse notation a bit so that when we say “hyperplane,” we will always mean an element of $(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ unless otherwise stated.

Separating hyperplanes

Definition 8

Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ denote a non-empty set \mathcal{X} of vectors in \mathbb{R}^n that are separated into the two (disjoint) non-empty sets \mathcal{X}_+ and \mathcal{X}_- . Such a collection of sets is called a training data set. A hyperplane $H \subseteq \mathbb{R}^n$, described by $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$, separates $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ iff

$$\langle \vec{w}, \vec{x}_+ \rangle - c > 0 \quad \& \quad \langle \vec{w}, \vec{x}_- \rangle - c < 0$$

for all $\vec{x}_+ \in \mathcal{X}_+$ and for all $\vec{x}_- \in \mathcal{X}_-$. In this case, H is said to be a separating hyperplane for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. H marginally separates $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ iff

$$\langle \vec{w}, \vec{x}_+ \rangle - c \geq 1 \quad \& \quad \langle \vec{w}, \vec{x}_- \rangle - c \leq -1$$

for all $\vec{x}_+ \in \mathcal{X}_+$ and for all $\vec{x}_- \in \mathcal{X}_-$.

Support vector machine

Definition 9

Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ denote a training data set in \mathbb{R}^n . Let $\mathcal{S}_{\mathcal{X}} \subseteq (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ denote the set of hyperplanes that marginally separate $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. Let $f : \mathcal{S}_{\mathcal{X}} \rightarrow \mathbb{R}$ be the function defined by

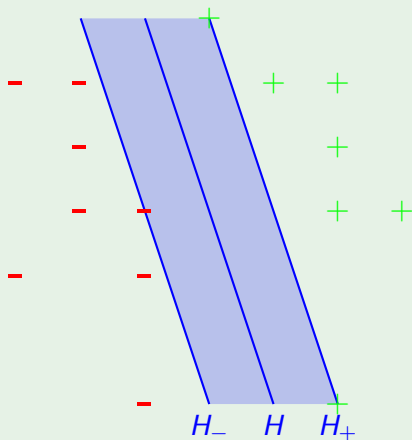
$$(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R} \ni (\vec{w}, c) \mapsto f(\vec{w}, c) := \frac{2}{\|\vec{w}\|},$$

i.e. the margin. A support vector machine (SVM) for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ is a maximum of f , i.e. an SVM is a marginally separating hyperplane $(\vec{w}, c) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ such that $\frac{1}{\|\vec{w}'\|} \leq \frac{1}{\|\vec{w}\|}$ for every other marginally separating hyperplane $(\vec{w}', c') \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$.

Support vector machine

Example 10

An SVM is a hyperplane that maximizes the margin



The support vectors of a marginally separating hyperplane

How will we know when we have found an SVM? The key to determining marginally separating hyperplanes will be in the possible support vectors. Our goal will then be to maximize the margin over all possible support vectors.

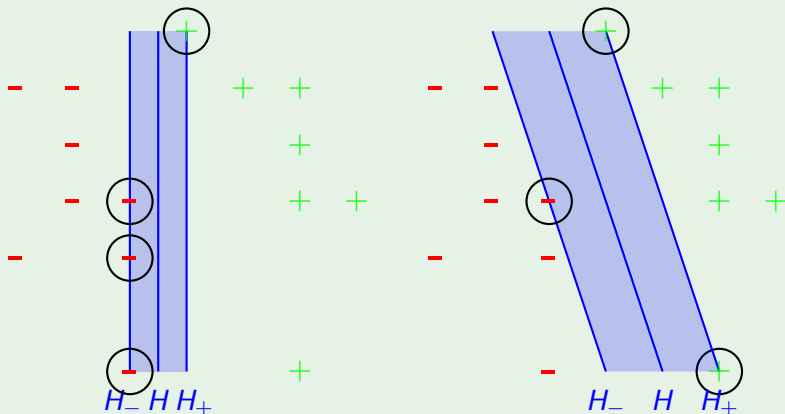
Definition 11

Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be a training data set and let (\vec{w}, c) be a marginally separating hyperplane for this set. The elements of $\mathcal{X} \cap H_{\pm}$ are called support vectors for (\vec{w}, c) . The set of support vectors is denoted by $H_{\mathcal{X}}^{\text{supp}}$. The notation $H_{\mathcal{X}_{\pm}}^{\text{supp}} := H_{\mathcal{X}}^{\text{supp}} \cap H_{\pm}$ will also be used to denote the set of positive and negative support vectors.

Examples of support vectors

Example 12

Two examples of support vectors have been circled for two different separating marginal hyperplanes for the same training data set.



“Enlarging the margins” Lemma Statement

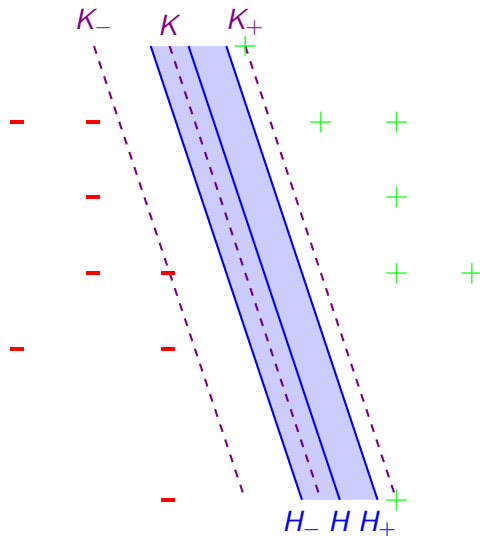
Lemma 13

Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be a training data set and let (\vec{w}, c) be a marginally separating hyperplane for this set. Then there exists a unique marginally separating hyperplane (\vec{v}, d) such that

$$\hat{v} = \hat{w} \quad \& \quad \frac{2}{\|\vec{v}\|} = \min_{\substack{\vec{x}_+ \in \mathcal{X}_+ \\ \vec{x}_- \in \mathcal{X}_-}} \langle \vec{x}_+ - \vec{x}_-, \hat{w} \rangle.$$

In other words, if the marginal planes do not contain any of the training data set, then the separating hyperplane can be translated and the margin width can be enlarged until the margin touches both positive and negative training data sets. We use the minimum orthogonal distance between the positive and negative training data sets to define the new margin width.

“Enlarging the margins” Lemma Picture



“Enlarging the margins” Lemma Proof

Proof.

Notice that $\left(\frac{c-1}{\|\vec{w}\|}\right) \hat{w} \in H_-$, $\left(\frac{c}{\|\vec{w}\|}\right) \hat{w} \in H$, and $\left(\frac{c+1}{\|\vec{w}\|}\right) \hat{w} \in H_+$ provide explicit vectors in each of these three hyperplanes. Set m_+ to be the remaining minimum orthogonal distance between H_+ and \mathcal{X}_+ and set m_- to be the remaining minimum orthogonal distance between H_- and \mathcal{X}_- , namely

$$m_{\pm} := \min_{\vec{x}_{\pm} \in \mathcal{X}_{\pm}} \theta(\vec{x}_{\pm}) \left(\langle \vec{x}_{\pm}, \hat{w} \rangle - \left(\frac{c \pm 1}{\|\vec{w}\|} \right) \right),$$

where $\theta : \mathcal{X} \rightarrow \mathbb{R}$ is the function defined by

$$\mathcal{X} \ni \vec{x} \mapsto \theta(\vec{x}) := \begin{cases} +1 & \text{if } \vec{x} \in \mathcal{X}_+ \\ -1 & \text{if } \vec{x} \in \mathcal{X}_- \end{cases}.$$

“Enlarging the margins” Lemma Proof

Proof.

Therefore, the planes K_{\pm} containing the vectors $\left(\frac{c \pm 1}{\|\vec{w}\|} \pm m_{\pm}\right) \hat{w}$ that are perpendicular to \hat{w} intersect \mathcal{X}_{\pm} but do not contain points of \mathcal{X} on the interior of their margin. By Lemma 6, there exists a $(\vec{v}, d) \in (\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R}$ that describes these marginal separating hyperplanes. In fact, they can be given explicitly by

$$\vec{v} := \left(\frac{2}{\frac{2}{\|\vec{w}\|} + m_+ + m_-} \right) \hat{w}$$

and

$$d := \left\langle \vec{v}, \left(\frac{c + 1}{\|\vec{w}\|} + m_+ \right) \hat{w} \right\rangle - 1 = \frac{2c + \|\vec{w}\|(m_+ - m_-)}{2 + \|\vec{w}\|(m_+ + m_-)}.$$



Existence of SVMs

Theorem 14

Let $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$ be training data set for which there exists a separating hyperplane for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$. Then there exists a unique SVM for $(\mathcal{X}, \mathcal{X}_+, \mathcal{X}_-)$.

Proof.

By Lemma 13, it suffices to maximize the margin function f on the subset $\mathcal{S}_{\mathcal{X}}^{\text{supp}} \subseteq \mathcal{S}_{\mathcal{X}}$ consisting of marginally separating hyperplanes that have both positive and negative support vectors, namely on

$$\mathcal{S}_{\mathcal{X}}^{\text{supp}} := \{(\vec{w}, c) \in \mathcal{S}_{\mathcal{X}} : H_{\pm} \cap \mathcal{X}_{\pm} \neq \emptyset\}.$$

The goal is therefore to maximize the margin function, which is a function of (\vec{w}, c) , subject to the constraint

$$\theta(\vec{x})(\langle \vec{w}, \vec{x} \rangle - c) - 1 = 0 \quad \forall \vec{x} \in \mathcal{S}_{\mathcal{X}}^{\text{supp}}.$$

Proof of existence of SVMs

Proof.

Maximizing the margin function is equivalent to minimizing the function

$$(\mathbb{R}^n \setminus \{\vec{0}\}) \times \mathbb{R} \ni (\vec{w}, c) \xrightarrow{g} \frac{1}{2} \|\vec{w}\|^2 - \sum_{\vec{x} \in \mathcal{X}} \alpha_{\vec{x}} \left(\theta(\vec{x}) (\langle \vec{w}, \vec{x} \rangle - c) - 1 \right).$$

Here, $\alpha_{\vec{x}} = 0$ for all $\vec{x} \in \mathcal{X} \setminus H_{\mathcal{X}}^{\text{supp}}$ (remember, $H_{\mathcal{X}}^{\text{supp}}$ is the set of support vectors for the given hyperplane) and $\alpha_{\vec{x}}$ needs to be determined for all $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$. This condition guarantees that the function g equals $\frac{2}{f^2}$ when restricted to $\mathcal{S}_{\mathcal{X}}^{\text{supp}}$ (but notice that it does not equal $\frac{2}{f^2}$ on the larger domain $\mathcal{S}_{\mathcal{X}}$ of all marginally separating hyperplanes). The $\alpha_{\vec{x}}$ are called Lagrange multipliers. The extrema of g occur at points (\vec{w}, c) for which the derivative of g vanishes with respect to these coordinates

$$\left. \frac{\partial g}{\partial \vec{w}} \right|_{(\vec{w}, c)} = 0 \quad \& \quad \left. \frac{\partial g}{\partial c} \right|_{(\vec{w}, c)} = 0.$$

Proof of existence of SVMs

Proof.

The first of these two equations gives

$$\vec{w} = \sum_{\vec{x} \in \mathcal{X}} \alpha_{\vec{x}} \theta(\vec{x}) \vec{x}, \quad (1)$$

while the second equation gives

$$\sum_{\vec{x} \in \mathcal{X}} \alpha_{\vec{x}} \theta(\vec{x}) = 0, \quad (2)$$

which is a condition that the Lagrange multipliers have to satisfy. Plugging in these results back into the function g gives

$$g(\vec{w}, c) = \sum_{\vec{x} \in \mathcal{X}} \alpha_{\vec{x}} - \frac{1}{2} \sum_{\vec{x}, \vec{y} \in \mathcal{X}} \alpha_{\vec{x}} \alpha_{\vec{y}} \theta(\vec{x}) \theta(\vec{y}) \langle \vec{x}, \vec{y} \rangle$$

Proof of existence of SVMs

Proof.

Setting $\left. \frac{\partial g}{\partial \alpha_{\vec{x}}} \right|_{(\vec{w}, c)} = 0$ for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$ will give additional conditions that the Lagrange multipliers have to satisfy, which are given by

$$\theta(\vec{x})(\langle \vec{w}, \vec{x} \rangle - c) = 1$$

for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$. Plugging in the earlier result (1) for \vec{w} into this constraint gives

$$\sum_{\vec{y} \in \mathcal{X}} \alpha_{\vec{y}} \theta(\vec{y}) \langle \vec{y}, \vec{x} \rangle - \theta(\vec{x}) = c \quad (3)$$

for each $\vec{x} \in H_{\mathcal{X}}^{\text{supp}}$. However, there is one subtle point, and that is that we do not know what $H_{\mathcal{X}}^{\text{supp}}$ is. Nevertheless, there is still an optimization procedure left over, and it is based on the different possible choices of $H_{\mathcal{X}}^{\text{supp}}$, i.e. for each pair of nonempty subsets $X_{\pm} \subseteq \mathcal{X}_{\pm}$.

Proof of existence of SVMs

Proof.

For each choice of $H_{\mathcal{X}}^{\text{supp}}$, one has the linear system

$$\sum_{\vec{y} \in \mathcal{X}} \theta(\vec{y}) \alpha_{\vec{y}} = 0$$

$$\left\{ \sum_{\vec{y} \in \mathcal{X}} \theta(\vec{y}) \langle \vec{y}, \vec{x} \rangle \alpha_{\vec{y}} - c = \theta(\vec{x}) \right\}_{\vec{x} \in X_+ \cup X_-} \quad (4)$$

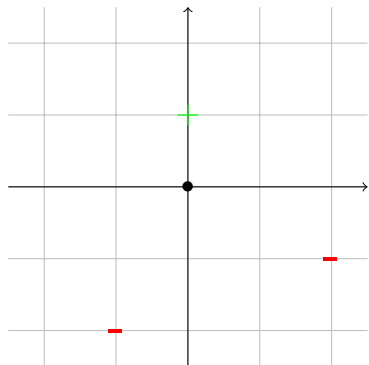
in the variables $\bigcup_{\vec{x} \in X_+ \cup X_-} \{\alpha_{\vec{x}}\} \cup \{c\}$ obtained from equations (2) and (3).

This describes a linear system of $|H_{\mathcal{X}}^{\text{supp}}| + 1$ equations in $|H_{\mathcal{X}}^{\text{supp}}| + 1$ variables. At least one of these systems is consistent because we have assumed that the training data set can be separated. Hence, a solution to the SVM problem exists since it is given by maximizing $\frac{1}{\|\vec{w}\|}$ over the (finite) set of pairs of nonempty subsets $X_{\pm} \subseteq \mathcal{X}_{\pm}$. □

Example of an SVM

Consider the training data set given by

$$\mathcal{X}_+ := \left\{ \vec{x}_+ := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad \& \quad \mathcal{X}_- := \left\{ \vec{x}_-^1 := \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \vec{x}_-^2 := \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}.$$



The inner products are given by

$$\langle \vec{x}_+, \vec{x}_+ \rangle = 1$$

$$\langle \vec{x}_+, \vec{x}_-^1 \rangle = -2$$

$$\langle \vec{x}_+, \vec{x}_-^2 \rangle = -1$$

$$\langle \vec{x}_-^1, \vec{x}_-^1 \rangle = 5$$

$$\langle \vec{x}_-^1, \vec{x}_-^2 \rangle = 0$$

$$\langle \vec{x}_-^2, \vec{x}_-^2 \rangle = 5$$

Example: case i) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-^1, \vec{x}_-^2\}$

For this case, (4) reads

$$\begin{array}{rcccc} \alpha_{\vec{x}_+} & -\alpha_{\vec{x}_-^1} & -\alpha_{\vec{x}_-^1} & & = & 0 \\ \alpha_{\vec{x}_+} & +2\alpha_{\vec{x}_-^1} & +\alpha_{\vec{x}_-^1} & -c & = & 1 \\ -2\alpha_{\vec{x}_+} & -5\alpha_{\vec{x}_-^1} & & -c & = & -1 \\ -\alpha_{\vec{x}_+} & & -5\alpha_{\vec{x}_-^1} & -c & = & -1 \end{array}$$

and the solution to this linear system is

$$\alpha_{\vec{x}_+} = \frac{5}{16}, \quad \alpha_{\vec{x}_-^1} = \frac{1}{8}, \quad \alpha_{\vec{x}_-^2} = \frac{3}{16}, \quad c = -\frac{1}{4}.$$

The associated vector \vec{w} and margin are therefore given by

$$\vec{w} = \frac{5}{16} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{8} \begin{bmatrix} -1 \\ -2 \end{bmatrix} - \frac{3}{16} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -1 \\ 3 \end{bmatrix} \implies \frac{2}{\|\vec{w}\|} = \frac{8}{\sqrt{10}}.$$

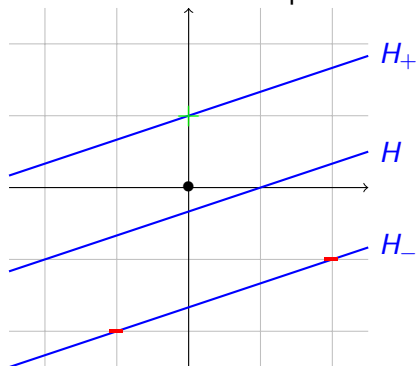
Example: case i) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-^1, \vec{x}_-^2\}$

The hyperplanes are then given by the solutions to the three equations

$$H : \frac{1}{4}(-x + 3y) + \frac{1}{4} = 0$$

$$H_+ : \frac{1}{4}(-x + 3y) + \frac{1}{4} = 1$$

$$H_- : \frac{1}{4}(-x + 3y) + \frac{1}{4} = -1$$



Example: case ii) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-^1\}$

For this case, $\alpha_{\vec{x}_-^2} = 0$ and (4) reads

$$\begin{array}{rcl} \alpha_{\vec{x}_+} & -\alpha_{\vec{x}_-^1} & = 0 \\ \alpha_{\vec{x}_+} & +2\alpha_{\vec{x}_-^1} & -c = 1 \\ -2\alpha_{\vec{x}_+} & -5\alpha_{\vec{x}_-^1} & -c = -1 \end{array}$$

and the solution to this linear system is

$$\alpha_{\vec{x}_+} = \frac{1}{5}, \quad \alpha_{\vec{x}_-^1} = \frac{1}{5}, \quad c = -\frac{2}{5}.$$

The associated vector \vec{w} and margin are therefore given by

$$\vec{w} = \frac{1}{5} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} -1 \\ -2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \implies \frac{2}{\|\vec{w}\|} = \frac{10}{\sqrt{10}}.$$

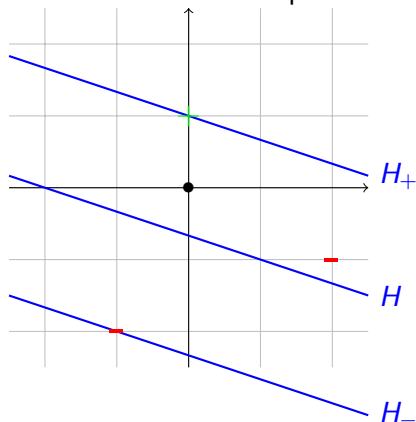
Example: case ii) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-^1\}$

The hyperplanes are then given by the solutions to the three equations

$$H : \frac{1}{5}(x + 3y) + \frac{2}{5} = 0$$

$$H_+ : \frac{1}{5}(x + 3y) + \frac{2}{5} = 1$$

$$H_- : \frac{1}{5}(x + 3y) + \frac{2}{5} = -1$$



Even though the margin is larger here, we must reject this solution because it is not a marginally separating hyperplane for the training data set.

Example: case iii) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-\}$

For this case, $\alpha_{\vec{x}_-^1} = 0$ and (4) reads

$$\begin{array}{rcl} \alpha_{\vec{x}_+} & -\alpha_{\vec{x}_-^1} & = 0 \\ \alpha_{\vec{x}_+} & +\alpha_{\vec{x}_-^1} & -c = 1 \\ -\alpha_{\vec{x}_+} & -5\alpha_{\vec{x}_-^1} & -c = -1 \end{array}$$

and the solution to this linear system is

$$\alpha_{\vec{x}_+} = \frac{1}{4}, \quad \alpha_{\vec{x}_-^2} = \frac{1}{4}, \quad c = -\frac{1}{2}.$$

The associated vector \vec{w} and margin are therefore given by

$$\vec{w} = \frac{1}{4} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \implies \frac{2}{\|\vec{w}\|} = \frac{2}{\sqrt{2}} = \frac{2\sqrt{5}}{\sqrt{10}}.$$

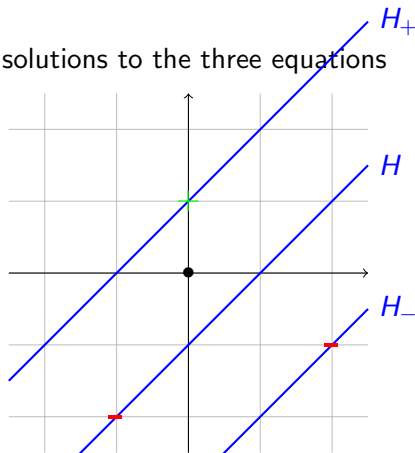
Example: case iii) $H_{\mathcal{X}}^{\text{supp}} = \{\vec{x}_+, \vec{x}_-^2\}$

The hyperplanes are then given by the solutions to the three equations

$$H : \frac{1}{5}(x + 3y) + \frac{2}{5} = 0$$

$$H_+ : \frac{1}{5}(x + 3y) + \frac{2}{5} = 1$$

$$H_- : \frac{1}{5}(x + 3y) + \frac{2}{5} = -1$$



We must also reject this solution because it is not a marginally separating hyperplane for the training data set. Hence, our first (\vec{w}, c) is the SVM.

Thank you!

Thanks for your attention!

References:

- 1 Alice Zhao, *Support Vector Machines: A Visual Explanation with Sample Python Code* (2017). Available at <https://youtu.be/N1v0golbjSc>.
- 2 Patrick Winston. 6.034 Artificial Intelligence “Lecture 16. Learning: Support Vector Machines.” Fall 2010. Massachusetts Institute of Technology: MIT OpenCourseWare, <https://ocw.mit.edu>. License: Creative Commons BY-NC-SA. Available at https://youtu.be/_PwhiWxHK8o.
- 3 Benjamin Russo. Private Communication.
- 4 Images of muffin from wikipedia (<https://en.wikipedia.org/wiki/Muffin>)
- 5 Images of cupcake from <https://littlecupcakes.com.au/product/red-velvet/>.