# An Application of Hahn-Banach theorem in Separation of Two Disjoint Point Sets in Two Dimensional Space

February 28, 2020

**Abstract**

In this article we give another approach of linear separation of two disjoint data sets instead of SVM (supporter vector machine). We use famous Hahn-Banach theorem and its proposition, strict separation theorem to derive a linear separator in two dimensional case. The construction of linear separator can be done by our precise analysis of Minkowski functional of convex hull, which is derived from Minkowski difference of two convex hulls of data sets. Our method gives a deterministic approach to linear separator, instead of iterative method used in SVM.

## 1 Introduction

Hahn-Banach theorem is a well known mathematical tool for functional analysis, In a word, it states that given a sublinear functional $p$ on space $E$, and a linear functional $\ell$ on subspace $V$, which satisfies $\ell \leq p$, we can extend the functional $\ell$ to linear functional $\tilde{\ell}$ on whole space $E$ and still satisfies the inequality. A direct consequence is the following: Given two disjoint sets $A$ and $B$, chosen any points $a_0 \in A$ and $b_0 \in B$, we construct set $C$:

$$C = \{y_0 : y_0 = a - b + b_0 - a_0, a \in A, b \in B\}$$

Considering Minkowski functional $p_C$ of set $C$, it guarantees that $p_C(x) \leq 1$ for $x \in C$ and $p_C(x) > 1$ for $x \notin C$. If we let $\ell(b_0 - a_0) = p_C(b_0 - a_0)$ and extend $\ell$ on whole space $\mathbb{R}^d$ to $\tilde{\ell}$ with restriction $\ell \leq p_C$, the following inequality will hold:

$$\max_{a \in A} \tilde{\ell}(a) \leq \min_{b \in B} \tilde{\ell}(b)$$

By computing the two sides of inequality, $\tilde{\ell}(x) = \gamma$ will separate two sets where:

$$\gamma \in [\max_{a \in A} \tilde{\ell}(a), \min_{b \in B} \tilde{\ell}(b)]$$

and it gives mathematical foundation of possibility of separating two disjoint sets by a hyperplane [1]. We will give details of the proof of above statements in section 3. However, there are some existence proof skills in proof of the theorem [9]. Thus we cannot derive the separating hyperplane directly from the proof.

Due to the work of Corinna Cortes and Vapnik, SVM gives a gradient approach to construct the separating hyperplane [4]. And it becomes the most popular mechanism for solving classification problem [2]. The current usage of sub-gradient method for SVM without kernel needs $O(ndT)$ time, for n examples, d dimensions, T steps [8]. The shortage of SVM is it does not guarantee we get the solution in a given time.

In this article, we review the proof of Hahn-Banach theorem, and in section 4 prove that the key functional, Minkowski functional is piecewise linear convex function by constructing the analytical form of Minkowski functional, when it is a functional of a convex polygon. The construction of analytical form cannot be done in general cases. Furthermore, we show that the minimum of Minkowski functional on a convex polygon can be computed directly once we have the convex hull, without using any gradient method.

We conclude that two disjoint data sets, separator in linear separation problem can be easily constructed using a convex polygon $K$ obtained from Minkowski difference of two data sets with the help of Minkowski functional. And the minimum of Minkowski functional makes construction steps possible and quickly be done in practice, instead of using any gradient method like SVM. In section 6, we give three results in 2D case. We show that for two disjoint sets, our separator can separate two sets accurately. And our separator behaves well even when two sets have a little of intersection. Our experiments is in two dimensional case, but the algorithm is expected to be used in any dimensional space.

From a higher perspective, this functional approach is using Minkowski functional of convex hull related to data set measures the whole data space.

# 2 Preliminaries

In the following sections, $Conv(A)$ denotes the convex hull of point set $A$. $A \setminus B$ denotes the set difference of A and B. $span\{x_0, x_1, \ldots, x_d\}$ denotes the set of all finite linear combinations of $x_0, x_1, \ldots, x_d$. $|\cdot|$ denotes 2-norm.

*Definition 2.1* (Minkowski difference). Given a vector space $V$ and two point sets $A$ and $B$ in $V$, the Minkowski difference $A \ominus B$ is defined as:

$$A \ominus B = \{a - b \mid a \in A \text{ and } b \in B\}$$

Without proof here, we use the fact that

$$Conv(A \ominus B) = Conv(A) \ominus Conv(B) \tag{1}$$

The proof of above equation is easy and can be found in [6].

*Definition 2.2* (sublinear functional). Given a vector space $V$ and a functional $P : V \to \mathbb{R}$, we call $P$ is a sublinear functional if:

$$\begin{cases} P(av) = aP(v), & \text{if } a \geq 0 \text{ and } v \in V \\ P(v_1 + v_2) \leq P(v_1) + P(v_2), & \text{if } v_1, v_2 \in V_0 \end{cases} \tag{2}$$

*Definition 2.3* (Minkowski functional). Given a vector space $V$ and a non empty set $K$ including 0, and 0 is not on the boundary of set $K$. For $v \in V$ define:

$$p_K(v) = \inf_{r>0}\{r : v/r \in K\}.$$

and functional $p_K$ completely characterized $K$ in that

$$p_K(v) \leq 1 \qquad \text{if and only if } v \in K$$

One important property of Minkowski functional is a sublinear functional [9].

*THEOREM 2.4* (Hahn-Banach Theorem). *Suppose $V_0$ is a linear subspace of $V$ and $p$ is a sublinear functional on $V$, and that we are given a linear functional $\ell$ on $V_0$ that satisfies*

$$\ell(v) \leq p(v), \qquad \textit{for all } v \in V_0$$

*Then $\ell$ can be extended to a linear functional $\tilde{\ell}$ on $V$ that satisfies*

$$\tilde{\ell}(v) \leq p(v), \qquad \textit{for all } v \in V$$

*THEOREM 2.5* (Strict Separation Theorem). *Given a vector space $V$, and two non empty convex set $A$ and $B$, which satisfy: $A \cap B = \emptyset$. Then there is a linear functional $\ell$, and a real number $\alpha \in \mathbb{R}$, s.t.*

$$\max_{a \in A} \ell(a) < \alpha < \min_{b \in B} \ell(b), \qquad a \in A$$

Since our construction of linear separator is based on the proof of theorem 2.4 and 2.5, we will show the proof of theorems in section 3 and analyze some functions in the proof in section 4.

# 3 Proofs

In this section we will prove theorem 2.4 and theorem 2.5. The procedure of proofs will be used in constructing linear separator.

## 3.1 Proof of Theorem 2.4

*Proof.* Given $\ell(x_0)$ for subspace $x_0 \in V_0$, we select any vector $x_1$ which is in space $V \setminus V_0$ and be ready to extend $\ell$ to the subspace $span\{x_0, x_1\}$. We can make a choice for the value of $\tilde{\ell}$ on $x_1$, so as to satisfy $\tilde{\ell}(x) \leq p(x), \forall x \in V$ if

$$a\tilde{\ell}(x_1) + b\tilde{\ell}(x_0) = \tilde{\ell}(ax_1 + bx_0) \leq p(ax_1 + bx_0), \qquad \forall a, b \in \mathbb{R}$$

where $\tilde{\ell(x)} = \ell(x)$ for all $x \in V_0$. If $a = 0$, above inequality trivially holds. If $a \neq 0$

$$\begin{cases} \tilde{\ell}(x_1) \leq \dfrac{p(ax_1 + bx_0) - b\tilde{\ell}(x_0)}{a}, a > 0 \\ \tilde{\ell}(x_1) \geq \dfrac{b\tilde{\ell}(x_0) - p(ax_1 + bx_0)}{-a}, a < 0 \end{cases}$$

Then we get:

$$\tilde{\ell}(w) - p(-x_1 + w) \leq \tilde{\ell}(x_1) \leq p(x_1 + w') - \tilde{\ell}(w'), \qquad \forall w, w' \in V_0$$

Since

$$\begin{aligned} \tilde{\ell}(w) + \tilde{\ell}(w') =& \tilde{\ell}(w - x_1 + w' + x_1) \\ \leq& p(w - x_1 + w' + x_1) \\ \leq& p(-x_1 + w) + p(x_1 + w') \end{aligned}$$

always holds. The first inequality holds because of $w + w' \in V_0$, The second equality holds because of sublinear property of $p$. The Thus the upper bound of $\tilde{\ell}(x_1)$ is

$$\min_{w \in V_0} \{-\tilde{\ell}(w) + p(w + x_1)\} \tag{3}$$

and the lower bound is

$$\max_{w' \in V_0} \{\tilde{\ell}(w') - p(w' - x_1)\} \tag{4}$$

We can choose any real number between two bounds as value of $\tilde{\ell}(x_1)$. In practice, we can choose the mean of two bounds as value of $\tilde{\ell}(x_1)$. We can now extend subspace $V_0$ to $span\{x_0, x_1\}$ and letting

$$\tilde{\ell}(\lambda_0 x_0 + \lambda_1 x_1) = \lambda_0 \tilde{\ell}(x_0) + \lambda_1 \tilde{\ell}(x_1)$$

We keep doing above step until $V = V_0$ with $\tilde{\ell}(x)$ mapping $V$ to $\mathbb{R}$ □

For easier reading, we omit the difference between the symbol $\tilde{\ell}$ and $\ell$ in following sections. We will keep using symbol $\ell$ after extension.

## 3.2 Proof of Theorem 2.5

*Proof.* Choose $a_0 \in A$, $b_0 \in B$ which $b_0 - a_0$ is not on the boundary of $Conv(A \ominus B)$. Let $x_0 = b_0 - a_0$. Construct set $C$:

$$C = \{y_0 : y_0 = a - b + x_0, a \in A, b \in B\} \tag{5}$$

$C$ has following properties:

- $C$ is convex set, because of equation 1 and shifting set by $x_0$.
- $x_0 \notin C$ and $0 \in C$, because of $A \cap B = \emptyset$

- $\forall x \in C$, $p_C(x) \leq 1$, where $p_C$ is Minkowski functional of $C$.

Since the Minkowski functional $p_C$ is a sublinear functional, by Hahn-Banach Theorem, there exists a linear functional $\ell$, s.t.

$$\ell \leq p_C \text{ and } \ell(x_0) = p_C(x_0) > 1$$

$p_C(x_0) < \infty$ since $x_0$ is not on the boundary of $Conv(A \ominus B)$. Then we can extends functional $\ell(x_0)$ to space $V_0 = span\{x_0\}$ by letting $\ell(kx_0) = k\ell(x_0)$. For $k \geq 0$, the inequality $\ell \leq p_C$ holds by property (2). For $k < 0$, the inequality $\ell \leq p_C$ holds immediately.

For all $x \in C$, we have:

$$\ell(a - b + x_0) \leq p_C(a - b + x_0) < 1$$

We get:

$$\ell(a) < \ell(b) + 1 - \ell(x_0) < \ell(b), \qquad \forall a \in A, \forall b \in B \tag{6}$$

The last inequality holds because of $\ell(x_0) > 1$. Thus:

$$\max_{a\in A} \ell(a) < \min_{b\in B} \ell(b)$$

Thus there exists $\alpha \in \mathbb{R}$ which satisfies:

$$\ell(a) < \alpha < \ell(b), \qquad \forall a \in A, b \in B$$

$\square$

In practice, we can choose $\ell(x_0) = p_C(x_0)$, and extend $\ell$ as procedure of proof in section 3.1. After extending $\ell(x)$ to entire vector space $V$, we select

$$\gamma \in [\max_{a\in A} \ell(a), \min_{b\in B} \ell(b)]$$

for which hyperplane $\ell(x) = \gamma$ will separate set $A$ and $B$. In practice, we can choose

$$\gamma = \frac{\max_{a\in A} \ell(a) + \min_{b\in B} \ell(b)}{2} \tag{7}$$

as our optimal $\gamma$.

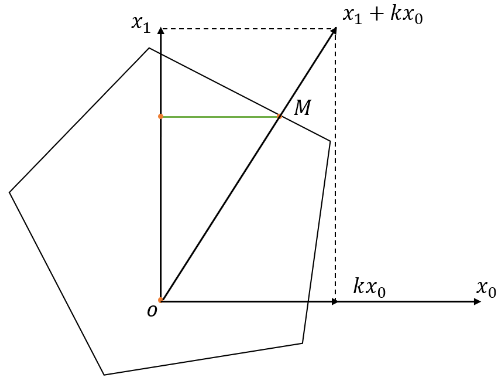
# 4 Analysis of two bounds of $\ell(x_1)$

In this section, we will show that if we choose $\ell(x_0) = p_C(x_0)$, the two bounds of $\ell(x_1)$ will be equal, where $C$ is defined as equation (5). Since upper bound (3) of $\ell(x_1)$ has the same form with lower bound (4) except for a minus sign, we only need to analyze the upper bound and use the same routine on lower bound.

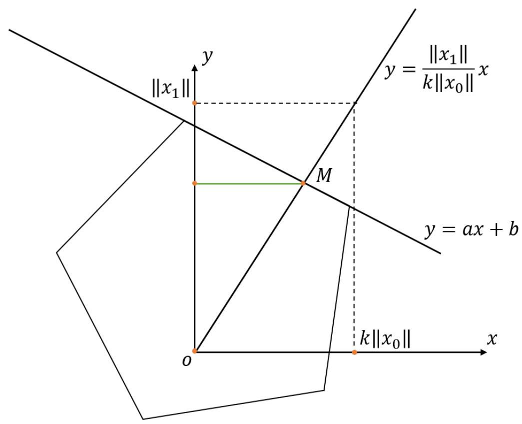
We rewrite the function $p_C(x_1 + w) - \ell(w)$ as

$$f(k) = p_C(x_1 + kx_0) - k\ell(x_0) \tag{8}$$

We first prove $p_C(x_1 + kx_0)$ is piecewise linear continuous convex function of $k$. Then we analyze function $f(k)$. We choose $x_1$ orthogonal to $x_0$ for the sake of easier analysis.




(a) The intersection of the polygon and the ray




(b) The coordinate system and related lines

Figure 1: The diagram of proof.

### 4.1 Analysis of $p_C(x_1 + kx_0)$

In this part, we will show that $p_C(x_1 + kx_0)$ is piecewise linear continuous function of $k$. We denote the point at which the ray in direction of $x_1 + kx_0$ intersects polygon as $M$. See (a) in figure 1. Then we set the x-axis in direction of $x_0$ and the y-axis in direction of $x_1$. The result coordinate system as (b) in figure 1. The coordinate value of $M$ satisfies

$$\begin{cases} y = ax + b \\ y = \dfrac{\|x_1\|}{k\|x_0\|}x \end{cases} \tag{9}$$

Solving equation (9) we get the x-axis component of $M$ is:

$$x_M = \frac{bk|x_0|}{-ak|x_0|+|x_1|}$$

By triangle similarity, $p_C(x_1 + kx_0) = \frac{k|x_0|}{x_M}$, we get:

$$p_C(x_1 + kx_0) = -\frac{a\,|x_0|}{b}k + \frac{|x_1|}{b} \tag{10}$$

Thus $p_C(x_1 + kx_0)$ is a linear function of $k$ on each segment and is a piecewise linear continuous function. And $p_C(k_1 + kx_0)$ is a convex function of $k$ since:

$$\begin{aligned}
&f(\lambda k_1 + (1-\lambda)k_2)\\
=&p_C(x_1 + (\lambda k_1 + (1-\lambda)k_2)x_0) - (\lambda k_1 + (1-\lambda)k_2)\ell(x_0)\\
=&p_C(\lambda(x_1 + kx_0) + (1-\lambda)(x_1 + k_2x_0))\\
&-(\lambda k_1 + (1-\lambda)k_2)\ell(x_0)\\
\leq&p_C(\lambda(x_1 + kx_0)) - \lambda k_1\ell(x_0)\\
&+p_C((1-\lambda)(x_1 + k_2x_0)) - (1-\lambda)k_2\ell(x_0)\\
=&\lambda(p_C(x_1 + kx_0) - \lambda k_1\ell(x_0))\\
&+(1-\lambda)(p_C(x_1 + k_2x_0) - k_2\ell(x_0))\\
=&\lambda f(k_1) + (1-\lambda)f(k_2)
\end{aligned}$$

The inequality and the following equality holds because of sublinear properties of $p_C$. Thus $f(k)$ is a piecewise linear continuous convex function.

## 4.2 The upper bound of $\ell(x_1)$

If we choose $\ell(x_0) = p_C(x_0)$, equation (9) for $x_0$ becomes:

$$\begin{cases} y = ax + b \\ y = 0 \end{cases} \tag{11}$$

we get:

$$p_C(kx_0) = -\frac{a\,|x_0|}{b}k \tag{12}$$

Substitute equation (10) and (12) for (8). we can get explicit form of $f(k)$:

$$f(k) = a^*k + b^* \tag{13}$$

where:

$$a^* = -\frac{a_s\,|x_0|}{b_s} + \frac{a_0\,|x_0|}{b_0} \tag{14}$$

and:

$$b^* = \frac{|x_1|}{b_s} \tag{15}$$

where $a_s$, $b_s$ is equation of segment which intersects the ray in direction of $x_1 + kx_0$, and $a_0$, $b_0$ is equation of segment which intersects the ray in direction of $x_0$.

We denote the point at which line $y = a_ix + b_i$ and x axis intersect as $d_i$, $d_i$ is

$$d_i = -\frac{b_i}{a_i}$$

By property of convex polygon, $d_s \geq d_0$ if $d_s > 0$, and $d_s < d_0$ if $d_s < 0$. Since $d_0 > 0$ we have:

$$\frac{1}{d_s} - \frac{1}{d_0} \leq 0$$

Thus $a^*$ is always non-positive and equal to zero if $x_1 + kx_0$ and $x_0$ intersect the same segment. Thus the minimum of $f(k)$ is

$$\ell(x_1) \leq \min_k f(k) = \frac{|x_1|}{b_0} \tag{16}$$

which is the upper bound of $\ell(x_1)$.

### 4.3 The lower bound of $\ell(x_1)$

The lower bound of $\ell(x_1)$ can be computed as negative of upper bound of $\ell(-x_1)$. The upper bound of $\ell(-x_1)$ is:

$$g(k) = \left(-\frac{a_s\,|x_0|}{b_s} + \frac{a_0\,|x_0|}{b_0}\right)k - \frac{|x_1|}{b_s} \tag{17}$$

Thus the lower bound of $\ell(x_1)$ is:

$$\ell(x_1) \geq \frac{|x_1|}{b_0} \tag{18}$$

By equation (16) and (18), we conclude that

$$\ell(x_1) = \frac{|x_1|}{b_0} \tag{19}$$

## 5 Construct of functional $\ell(x)$

In this section, we will construct the explicit form of $\ell(x)$ in $d$ dimensional space, the two dimensional form is the special case when $d = 2$.

After extending $\ell$ to whole space $V$. we obtain $d$ independent vectors $x_1, x_2, \ldots, x_d$ and their images $\ell(x_1), \ell(x_2), \ell(x_3), \ldots, \ell(x_d)$. Our input bases is standard basis and output basis is $x_1, x_2, \ldots, x_d$. Thus the $d \times d$ matrix of new basis is:

$$W = \begin{bmatrix} x_1 & x_2 & x_3 & \ldots & x_d \end{bmatrix}$$

The image column vectors is:

$$L = \begin{bmatrix} \ell(x_1) & \ell(x_2) & \ell(x_3) & \ldots & \ell(x_d) \end{bmatrix}^T$$

for any point $x$ we can decompose $x$ as:

$$x = a_1 x_1 + a_2 x_2 + \cdots + a_d x_d \tag{20}$$

and apply functional $\ell$ we have

$$\ell(x) = a_1 \ell(x_1) + a_2 \ell(x_2) + \cdots + a_d \ell(x_d) \tag{21}$$

Solve equation (20) and (21) and using matrix form:

$$\ell(x) = L^T W^{-1} x$$

By conclusion of subsection 3.2, we choose $\gamma$ as in equation 7. Thus hyperplane $\ell(x) = \gamma$ will separate two disjoint sets.

# 6 Experimental Results

In this section, we show three linear separation results. Result (a) and (b) in figure 2 are using bivariate gaussian distributions where:

$$\mu_A = \begin{bmatrix} -2 & 0 \end{bmatrix} \qquad \sigma_A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

and

$$\mu_B = \begin{bmatrix} 4 & 5 \end{bmatrix} \qquad \sigma_B = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$$

Result (a) shows our separator can separate two disjoint sets accurately. Result (b) shows even the data sets have a little intersection, our separator still behaves well. Result (c) comes from iris dataset in the python sklearn library.

# 7 Discussions

We notice that if $A$ and $B$ are disjoint set, then $Conv(A \ominus B)$ will not include origin. Conversely, if $A$ and $B$ are not disjoint, $Conv(A \ominus B)$ always contains origin. This observation holds even if we apply a nonlinear feature map on data set. It encourages us to find a new method to test if two sets are disjoint after applying a nonlinear feature map.

Considering analysis in section 4, a reasonable guess is in $d$ dimensional space the minimum of $f(k_0, k_1, \cdots, k_{d-1})$ has the similar form with equation (16), if we know the equation of hyperplane which the vector $x_0 + x_1 + \cdots + x_{d-2}$ crosses. Then our separator can be used in $d$ dimensional space.

Our code implementation computes $Conv(A \ominus B)$ by computing convex hull of $A \ominus B$. For computing convex hull in two dimensional space, the classical algorithm Graham's scan costs $O(n \lg n)$ time, and a better algorithm Jarvis's march costs $O(nh)$ time, where $h$ is the vertices of convex hull [3]. The prune-and-search algorithm of Kirkpatrick and Seidel uses $O(n \lg h)$ time [7].

However, the size of $A \ominus B$ is $O(nm)$ if the size of $A$ and $B$ are $O(n)$ and $O(m)$. If we compute $Conv(A \ominus B)$ from $Conv(A)$ and $Conv(B)$, the time complexity will decrease significantly. Guibas and Seidel give an algorithm that computing $Conv(A \ominus B)$ in $O(n_A + m_B + c)$ time in three dimensional case, where $n_A$ and $m_B$ are the number of vertices of $Conv(A)$ and $Conv(B)$, and $c$ is the number of vertices of $Conv(A \ominus B)$ [5]. A similar algorithm for higher dimensional space will help to decrease computing time in higher dimensional space.

# References


[1] D.P. Bertsekas. *Convex Optimization Theory*. Athena Scientific optimization and computation series. Athena Scientific, 2009. pp.67-80.

[2] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006. pp.325-338.

[3] T.H. Cormen. *Introduction to Algorithms, 3rd Edition:*. MIT Press, 2009. pp.1029-1038.

[4] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

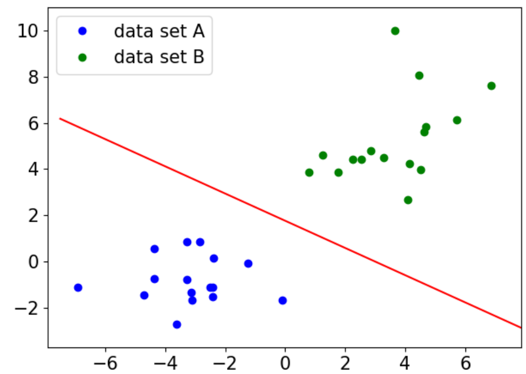
[5] L Guibas and R Seidel. Computing convolutions by reciprocal search. In *Proceedings of the Second Annual Symposium on Computational Geometry*, SCG '86, pages 90–99, New York, NY, USA, 1986. ACM.

[6] Post No Bulls (https://math.stackexchange.com/users/111742/post-no bulls). Prove convex hull of minkowski sum. Available at `https://math.stackexchange.com/q/625624` (2014/01/03).

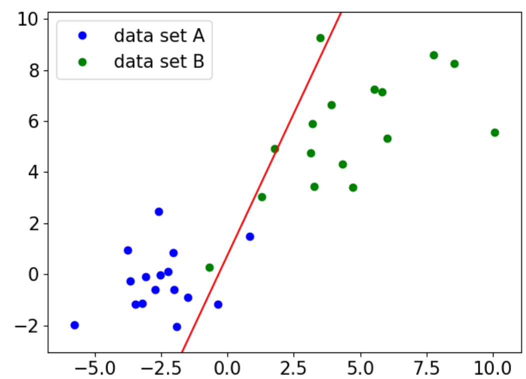
[7] David G Kirkpatrick and Raimund Seidel. The ultimate planar convex hull algorithm. *SIAM J. Comput.*, 15(1):287–299, February 1986.

[8] Lorenzo Rosasco. Class 05: Logistic regression and support vector machines. Available at `http://www.mit.edu/~9.520/fall18/slides/Class05_LogisticsSVM.pdf` (2018/09/19).

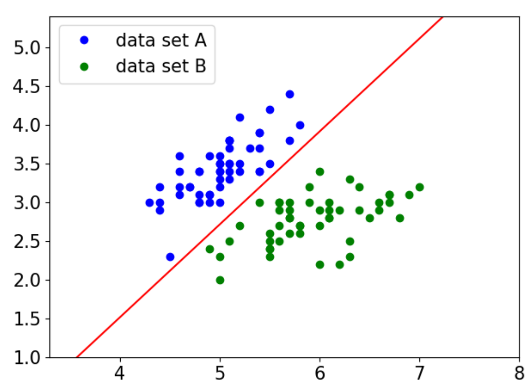
[9] E.M. Stein and R. Shakarchi. *Functional Analysis: Introduction to Further Topics in Analysis*. Princeton lectures in analysis. Princeton University Press, 2011. pp.16-21.




(a) two disjoint gaussian samples




(b) two gaussian samples with a little intersection




(c) datasets in python sklearn library

Figure 2: Three experimental results for linear separator.