Ch 9.2: Support Vector Classifier

Lecture 19 - CMSE 381

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Wed, 4/3, 2024

Announcements

Last time:

- 9.1 Maximal Margin Classifier

This lecture:

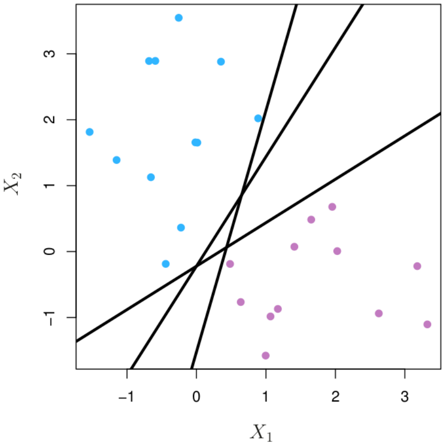
- 9.2 Support Vector Classifier

Announcements:

Section 1

Last time

Separating Hyperplane



Require that for every data point:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} > 0 \text{ if } y_i = 1$$
$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} < 0 \text{ if } y_i = -1$$

Equivalently

Require that for every data point

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) > 0$$

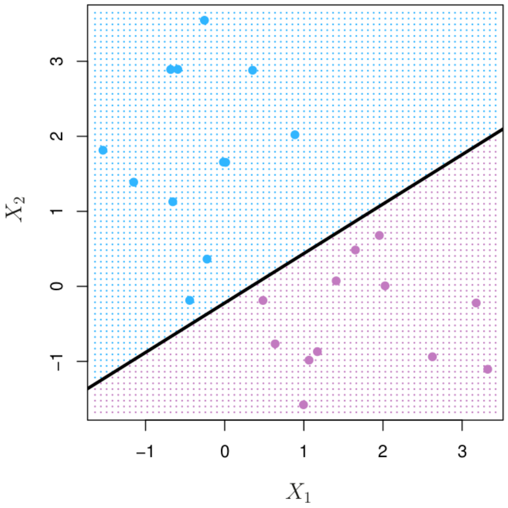
Separating hyperplane becomes a classifier

If you have a separating hyperplane:

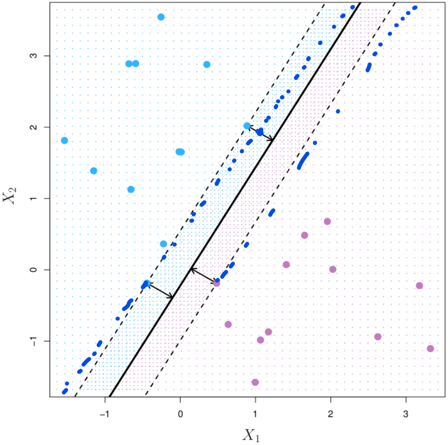
- Check

$$f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$$

- If positive, assign $\hat{y} = 1$
- If negative, assign $\hat{y} = -1$



Maximal margin classifier



- For a hyperplane, the *margin* is the smallest distance from any data point to the hyperplane.
- Observations that are closest are called *support vectors*.
- The *maximal margin hyperplane* is the hyperplane with the largest margin
- The classifier built from this hyperplane is the *maximal margin classifier*.

Mathematical Formulation

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,M}{\text{maximize}} \; M$$

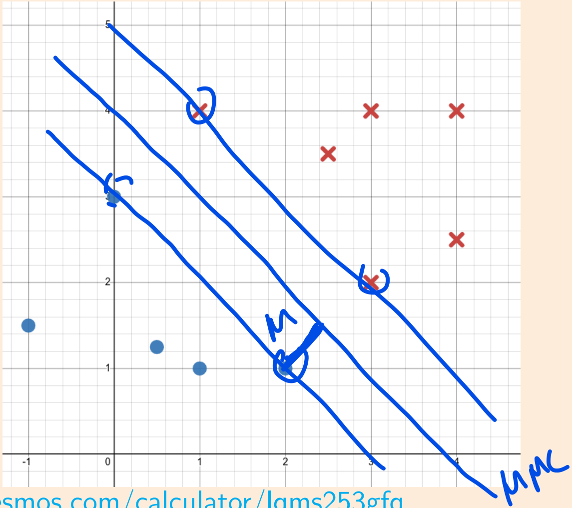
$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M \;\; \forall \, i = 1, \ldots, n$$

- Unit normal requirement, we can always write a given hyperplane this way.
- Last eq forces points to be on the correct side of the hyperplane in order for $M > 0$
- The product is the distance from the point to the hyperplane
- Making M as big as possible is the maximal margin hyperplane

Support vectors

Support vectors are those training data that either fall inside or on the boundary of the margin.



- Sketch the maximal margin hyperplane.
- What is the equation of this line in the form $\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$?
- Circle the support vectors. What is their distance from the line?
- Which side of the line has $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$

desmos.com/calculator/lqms253gfq

Example: Find support vectors

For $(0,3)$, $X_1 = 0$, $X_2 = 3$

$$\left|2\sqrt{2} - \frac{\sqrt{2}}{2}\cdot 0 - \frac{\sqrt{2}}{2}\cdot 3\right| = \frac{\sqrt{2}}{2} = M$$

so $(0,3)$ is on the boundary of the margin, so it is a support vector

Suppose the MMC of a dataset

1. is defined by the line $2\sqrt{2} - \frac{\sqrt{2}}{2}X_1 - \frac{\sqrt{2}}{2}X_2 = 0$ MMC
2. with the maximal margin equal to $\frac{\sqrt{2}}{2}$. $= M$

Which of the following training data are support vectors?

- $(0,3)$, $(3,2)$, $(4,2)$

$$y\cdot(\beta_0 + \beta_1 X_1 + \beta_2 X_2) > M \text{ or } = M$$

$$y\cdot\left(2\sqrt{2} - \frac{\sqrt{2}}{2}X_1 - \frac{\sqrt{2}}{2}X_2\right) > M \text{ or } = M$$

y only controls the sign, for distance, only need to focus on magnitude of LHS

For $(3,2)$, $X_1 = 3$, $X_2 = 2$

$$\left|2\sqrt{2} - \frac{\sqrt{2}}{2}\cdot 3 - \frac{\sqrt{2}}{2}\cdot 2\right| = \frac{\sqrt{2}}{2} = M$$

so it is also a support vector

For $(4,2)$,

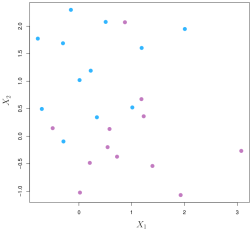
$$\left|2\sqrt{2} - \frac{\sqrt{2}}{2}\cdot 4 - \frac{\sqrt{2}}{2}\cdot 2\right| = \sqrt{2} > M$$

So it is not a support vector

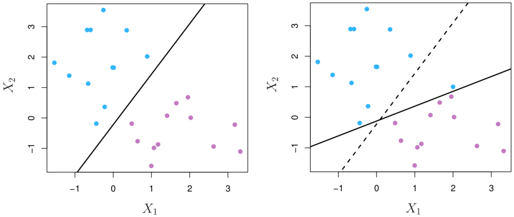
desmos.com/calculator/lqms253gfq

Problems

Might be no separating hyperplane



Sensitivity to new points



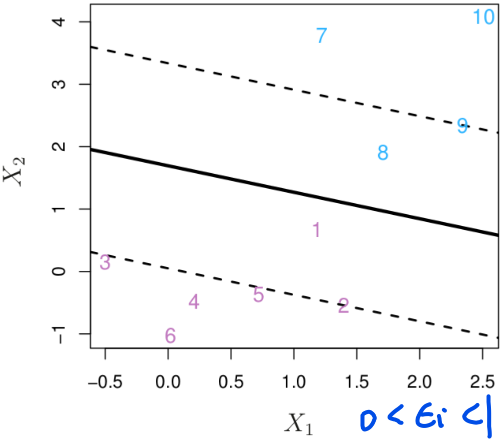
Section 2

Support Vector Classifier

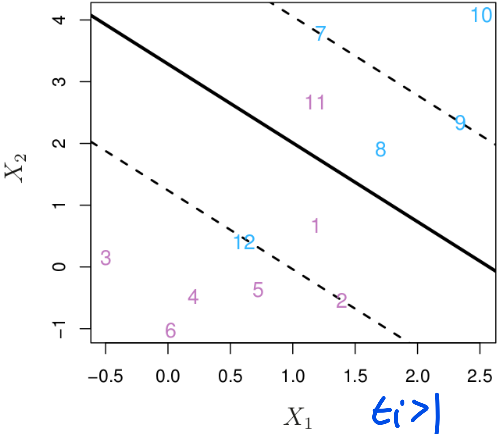
Basic idea

- Be ok with having a classifier that isn't quite perfect
- Aim for greater robustness to individual observations
- Better classification of most of the training observations
- Result is a support vector classifier
- Soft margin classifier

Soft margin



Some points on wrong side of margin



Some points on wrong side of hyperplane (Misclassified)

Mathematical Formulation of SVC

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,\, M}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M(1-\epsilon_i),$$

$$\underline{\epsilon_i \geq 0}, \quad \sum_{i=1}^{n} \epsilon_i \leq \boxed{C},$$

budget for misclassification

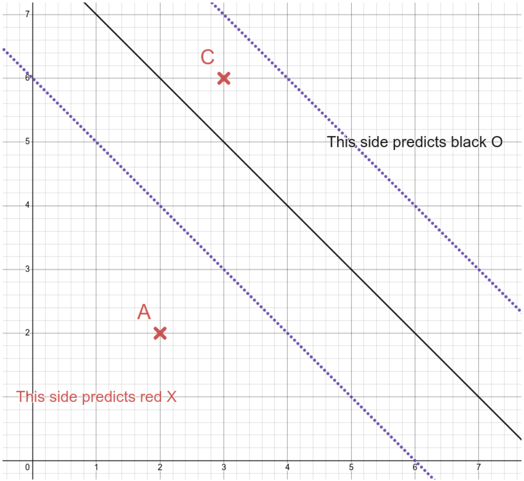
- C is nonnegative tuning parameter
- M is the width of the margin
- $\underline{\varepsilon_1, \cdots, \varepsilon_n}$ are slack variables allowing observations to go to the other side

$\epsilon_i = 1/2$ $M/2$

$\epsilon_i = 2$ $-M$

Find positive ε's that will satisfy this

Suppose I know $M = \sqrt{2}$ $\qquad y_i(-\frac{\cancel{4}\,6}{\sqrt{2}} + \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_1) \geq M(1-\varepsilon_i)$



- M is distance from center hyperplane to noted margin.
- A is on the correct side of the hyperplane.
  - Left side is already bigger than M, so set $\varepsilon_i = 0$.
- C is on the wrong side of the hyperplane for its label.
  - Left side of the equation is negative. What ε satisfies this?
  - $-\frac{\sqrt{2}}{2} = \sqrt{2}(1-\varepsilon)$
  - $\varepsilon = 3/2$

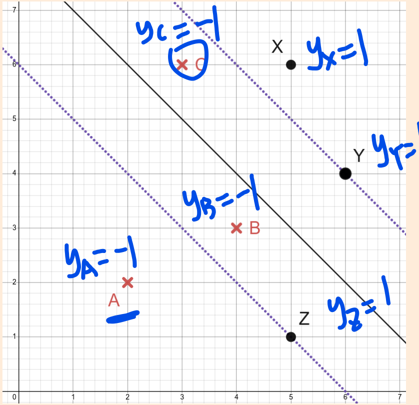
What is ε?

B: $(-1)\left(-\frac{8}{\sqrt{2}} + \frac{3}{\sqrt{2}} + \frac{6}{\sqrt{2}}\right) \geq \sqrt{2}(1-\varepsilon_i) \Leftrightarrow \frac{\sqrt{2}}{2} \geq \sqrt{2}(1-\varepsilon_i)$

Fix $M = \sqrt{2}$ $\quad y_i(-\frac{4}{\sqrt{2}} + \frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2) \geq M(1-\varepsilon_i)$ $\quad \Leftrightarrow \varepsilon_i \geq 1/2$

what is the smallest ε_i that make this constraint hold for each i ?

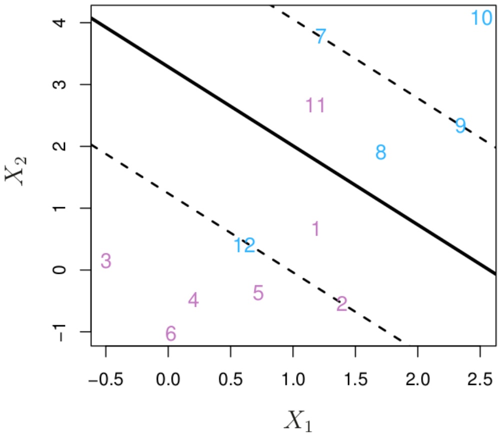
Fill in the table so that the inequality is satisfied.



Point	Left Side	ε_i	$M(1-\varepsilon_i)$
A	$2\sqrt{2}$	0	$\sqrt{2}$
B	$\sqrt{2}/2$	$1/2$	$\sqrt{2}/2$
C	$-\sqrt{2}/2$	$3/2$	$-\sqrt{2}/2$
X	$3/\sqrt{2}$	0	$\sqrt{2}$
Y	$\sqrt{2}$	0	$\sqrt{2}$
Z	$-\sqrt{2}$	2	$-\sqrt{2}$

What is ε?

- If $\varepsilon_i = 0$, then on correct side of margin
- If $\varepsilon_i > 0$ then on the wrong side of margin (Violated margin)
- If $\varepsilon_i > 1$ then on the wrong side of hyperplane



What is C?

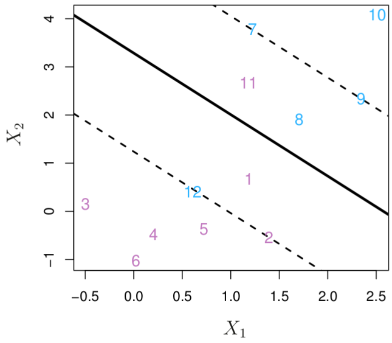
$$\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,\,M}{\text{maximize}} \quad M$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

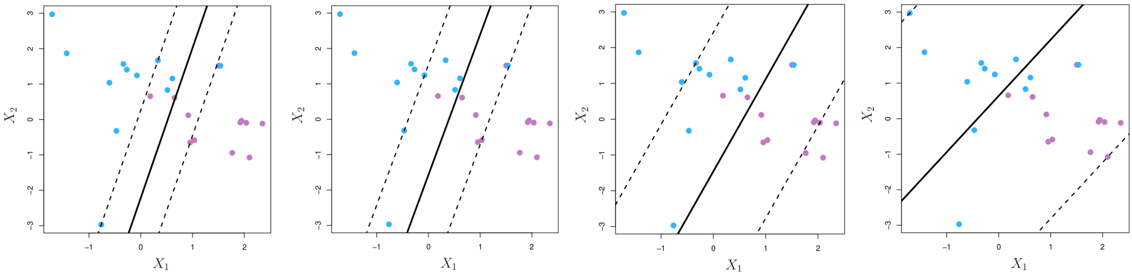
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C,$$

- Bounds sum of ε_i, so controls number & severity of violating margin (budget)
- $C = 0$ means no violations allowed
- $C > 0$ means at most C observations can be on wrong side of hyperplane
- In previous example, our total of ε was 4, so would be a valid hyperplane for C at most 4.



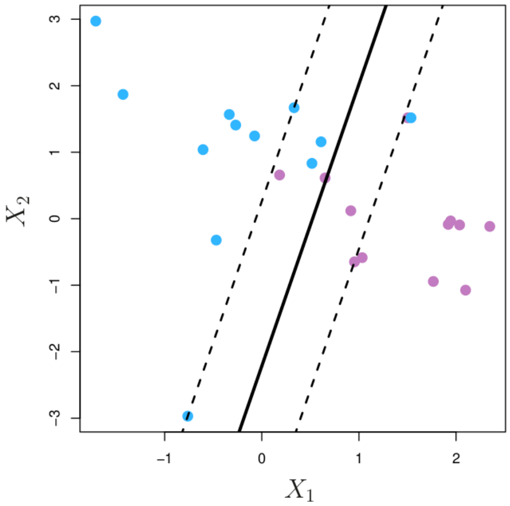
Examples messing with C



Increasing $C \rightarrow$

- For increasing C, we have more flexibility, so more points allowed violate margin/hyperplane
-

What affects the hyperplane?



- Only observations on the margin or violating the margin affect the hyperplane
- These observations are called support vectors
- Changing other points positions doesn't affect hyperplane found
-

SVC via inner products

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle$$

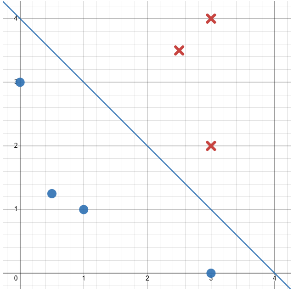
- Via some magic, there are coefficients α_i which give the linear support vector classifier
- In this notation, the x_i's are all the training points
- if x_i is not a support vector, then we can just set $\alpha_i = 0$.

- How to actually get it is outside the scope of this class
- To estimate the paramters $\alpha_1, \cdots, \alpha_n$ and β_0, need $\binom{n}{2} = n(n-1)/2$ inner products $\langle x_i, x_{i'} \rangle$

Example

$$-2\sqrt{2} + \frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2 = 0$$

$$-2\sqrt{2} + \frac{\sqrt{2}}{18}\langle (X_1, X_2), (0,3)\rangle + \frac{\sqrt{2}}{6}\langle (X_1, X_2), (3,2)\rangle = 0$$



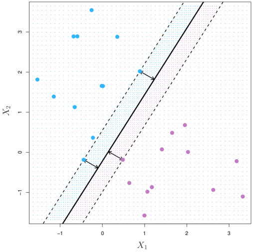
- $f(1,1) = -2\sqrt{2} + \frac{\sqrt{2}}{18}\langle (1,1), (0,3)\rangle + \frac{\sqrt{2}}{6}\langle (1,1), (3,2)\rangle$
- $= -2\sqrt{2} + \frac{\sqrt{2}}{18}\cdot 3 + \frac{\sqrt{2}}{6}\cdot 5$
- $= (-2 + \frac{3}{18} + \frac{5}{6})\sqrt{2} = -\sqrt{2}$

Maximal Margin Classifier

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,M}{\text{maximize}} \; M$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M \;\; \forall\, i = 1, \ldots, n$$



Support Vector Classifier

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,\, M}{\text{maximize}} \quad M$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M(1 - \epsilon_i),$$

$$\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C,$$