Ch 9.1: Maximum Margin Classifier

Lecture 18 - CMSE 381

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

Mon, April 1, 2024

Announcements

Last time:

- Ch 8: Random Forests

This lecture:

- Maximal Margin Classifier
- No jupyter notebook for this class

Announcements:

- Going over the exam on Wednesday

Section 1

Maximal Margin Classifier

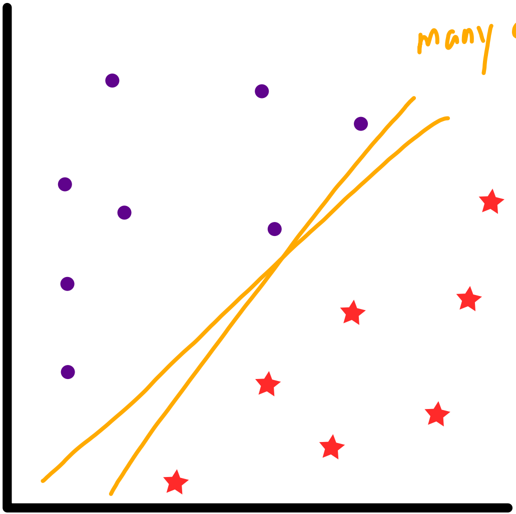
→ middle line ⇒ the classifier with the max margin

right margin

left margin

margin = min{ left margin, right margin }

The goal

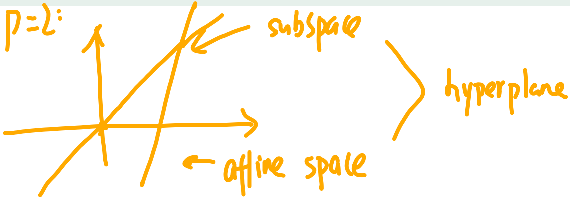


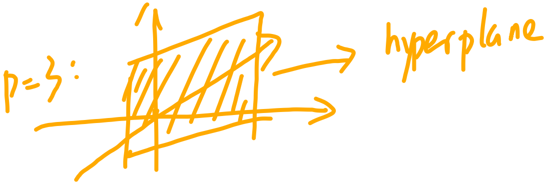
many seperating lines

- Today is all about classification
- Goal is to find a line in the middle that separates the two classes. Lots of options, but we want the one that stays as far from everyone as possible.
- For today, need to use $y \in \{-1, 1\}$ for this to make sense.

What is a hyperplane?

- In p-dimensional space, a hyperplane is a flat affine subspace of dimension $p - 1$
- Example for $p = 2, 3$



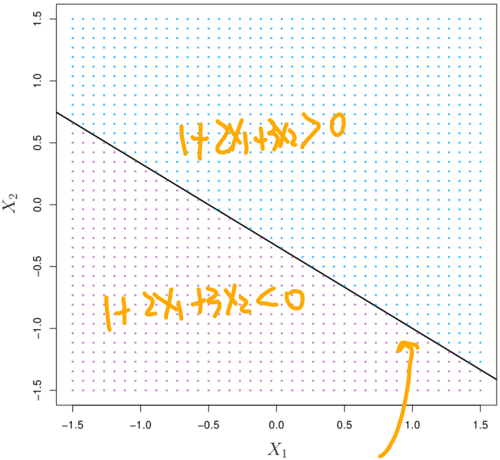


Mathematical definition of a hyperplane

$$H = \{ (X_1, \cdots, X_p) : \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0 \} \leftarrow p-1 \text{ dimensional}$$

- Being on the hyper plane means having a point $X = (X_1, X_2, \cdots, X_p)^T$ satisfying this
- Show that when $p = 2$, this is just a line

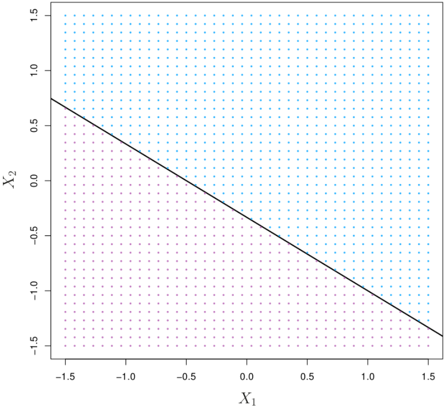
Hyperplane for $p = 2$



$1 + 2X_1 + 3X_2 = 0$

There are two sides to every hyperplane

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p < 0 \qquad \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p > 0$$



Classification Setup

Data matrix:

$$X = \begin{pmatrix} - & x_1^T & - \\ - & x_2^T & - \\ & \vdots & \\ - & x_n^T & - \end{pmatrix}_{n \times p}$$

$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \cdots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

Observations in one of two classes,
$y_i \in \{-1, 1\}$

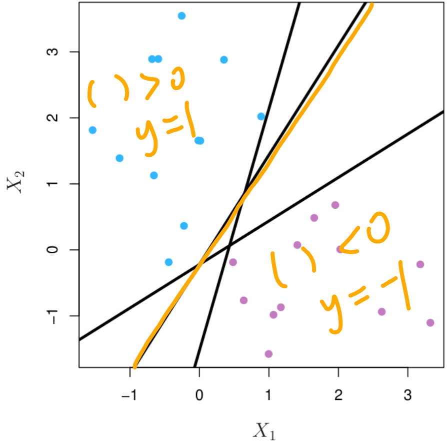
$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Separate out a test observation

$$x^* = (x_1^* \ \cdots \ x_p^*)^T$$

Separating Hyperplane

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



- Strict requirement
- No points can be on the wrong side of the line

Another way to say it

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

For all i:

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

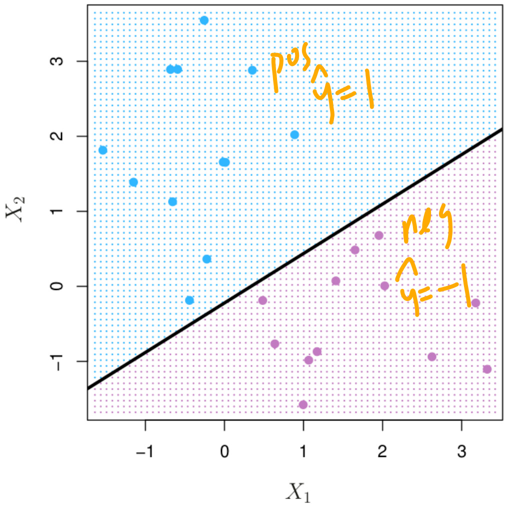
Separating hyperplane becomes a classifier

If you have a separating hyperplane:

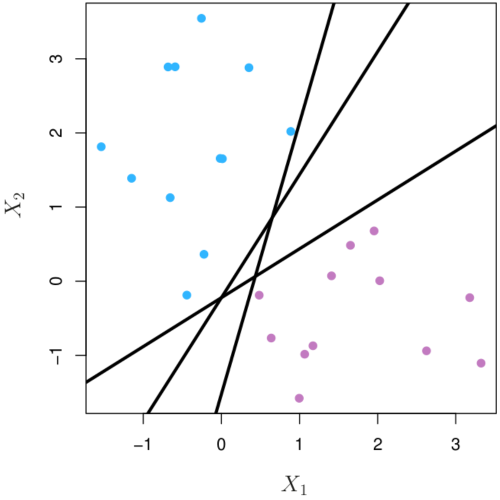
- Check

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

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

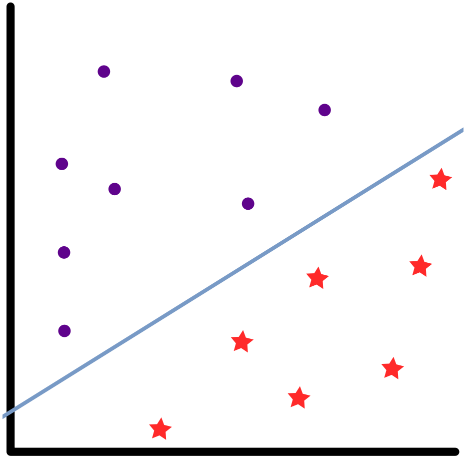


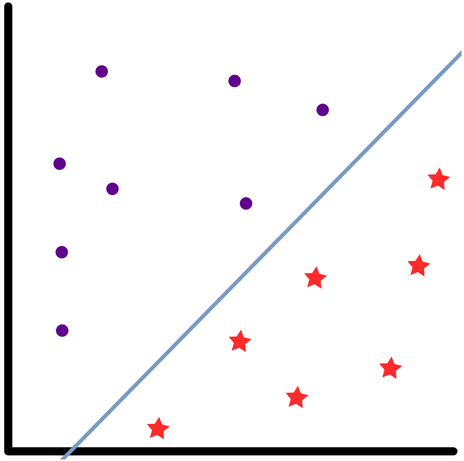
How do we pick?



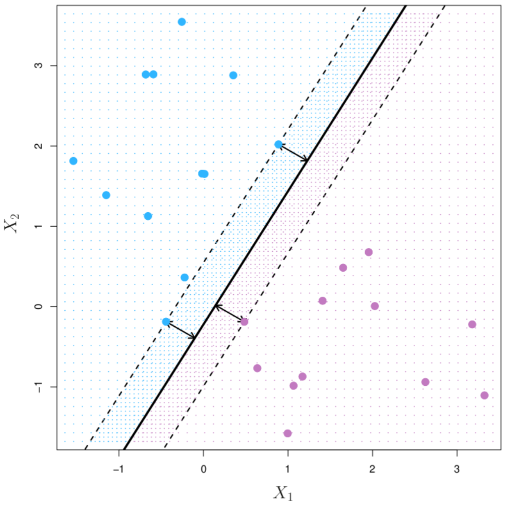
- *There's lots of separating hyperplanes here, so how do we find a "best" one?*
- Compute the perpendicular distance from each training observation to a given separating hyperplane
- Smallest such distance is the *margin*

Distance from an observation to a hyperplane

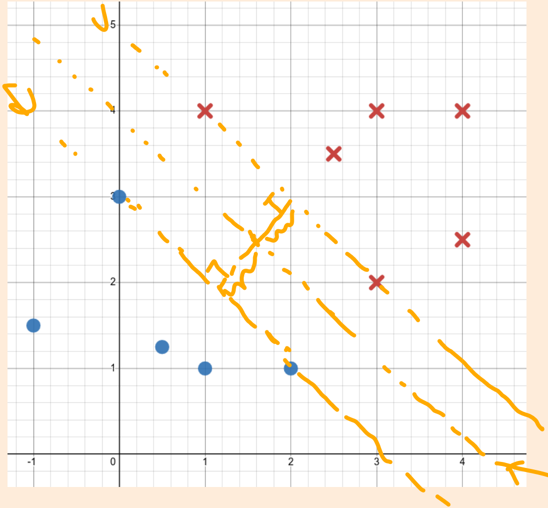




Maximal margin classifier



- *Maximal margin hyperplane* is the hyperplane with the largest minimum distance to the training observations
- Classify using it gives you the *maximal margin classifier*
- Observations that are closest are called *support vectors*
- If points moved slightly then MMH would move
- No effect if any other points moved



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

classification boundary

desmos.com/calculator/lqms253gfq

Mathematical Formulation

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

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

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

eg. $2X_1 + 3X_2 + 1 = 0$

$\beta = [\beta_1, \beta_2] = [2, 3]$

$\|\beta\| = \sqrt{4+9} = \sqrt{13}$

ensure uniqueness

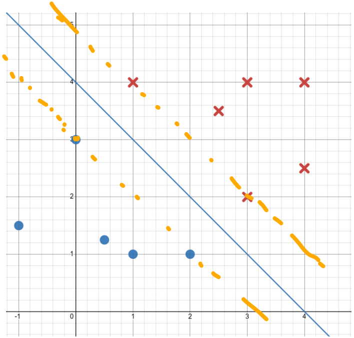
normalize

$\frac{2}{\sqrt{13}} X_1 + \frac{3}{\sqrt{13}} X_2 + \frac{1}{\sqrt{13}} = 0$

the first constraint is satisfied

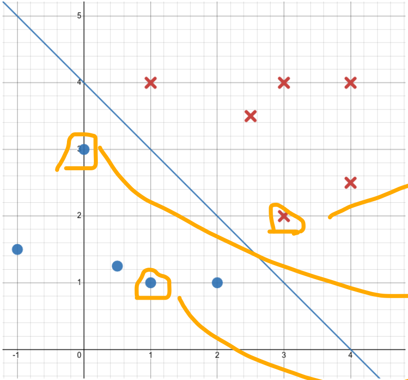
- Last eq forces points to be on the correct side of the hyperplane if $M > 0$
- Making M as big as possible is the maximal margin hyperplane

First constraint



- Had the equation $-4 + X_1 + X_2 = 0$
- Can divide by any number and get the same line.
- $\sum_{i=1}^{p} \beta_i^2 = 1^2 + 1^2 = 2$
- We get the same sloped line if we divide by $\sqrt{(\sum \beta_i)}$,
- Just need to solve for the intercept
- $-\frac{4}{\sqrt{2}} + \frac{1}{\sqrt{2}} X_1 + \frac{1}{\sqrt{2}} X_1 = 0$
- $-2\sqrt{2} + \frac{\sqrt{2}}{2} X_1 + \frac{\sqrt{2}}{2} X_1 = 0$
- Check that $\sum_{i=1}^{p} \beta_i^2$ is now 1
- Note that $y = 1$ is Red X, $y = -1$ is blue dots below the line

Second constraint

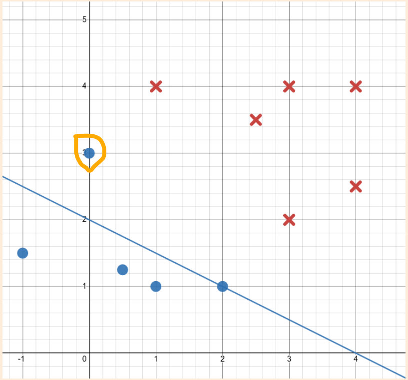


$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq M$$

- For point $(3, 2)$:
  - $y_i = 1$
  - $1(-2\sqrt{2} + \frac{\sqrt{2}}{2} \cdot 3 + \frac{\sqrt{2}}{2} \cdot 2) = \frac{\sqrt{2}}{2}$ $\geq M$
- For point $(0, 3)$:
  - $y_i = -1$
  - $-1(-2\sqrt{2} + \frac{\sqrt{2}}{2} \cdot 0 + \frac{\sqrt{2}}{2} \cdot 3) = \frac{\sqrt{2}}{2}$ $\geq M$
- For point $(1, 1)$
  - $y_i = -1$
  - $-1(-2\sqrt{2} + \frac{\sqrt{2}}{2} \cdot 1 + \frac{\sqrt{2}}{2} \cdot 1) = \sqrt{2}$

- Blue circles: $y_i = -1$
- Red Xs: $y_i = 1$
- $-2\sqrt{2} + \frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_1 = 0$

An example with a bad choice of hyperplane

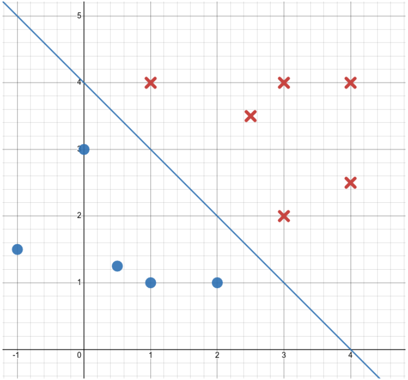


- Blue circles: $y_i = -1$
- Red Xs: $y_i = 1$
- $-\frac{4}{\sqrt{5}} + \frac{1}{\sqrt{5}}X_1 + \frac{2}{\sqrt{5}}X_1 = 0$

What is $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$ for the point $x_i = (0, 3)$?

- $-1(-\frac{4}{\sqrt{5}} + \frac{1}{\sqrt{5}} \cdot 0 + \frac{2}{\sqrt{5}} \cdot 3)$
- $-1(-4 + 6)\frac{1}{\sqrt{5}}$
- $\frac{-2}{\sqrt{5}}$ ← no M is smaller than this
- Note this is negative, so no positive M will make it so this is a valid line

Second constraint extra space



- Blue circles: $y_i = -1$
- Red Xs: $y_i = 1$
- $-2\sqrt{2} + \frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_1 = 0$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \geq M$$

- Ensures each on the correct side for a positive M
- When we have unit normal coefficients, the second constraint is distance from the input point to the diagonal.
- Making sure every distance is bigger than M and then make M as big as possible

Mathematical Formulation

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

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

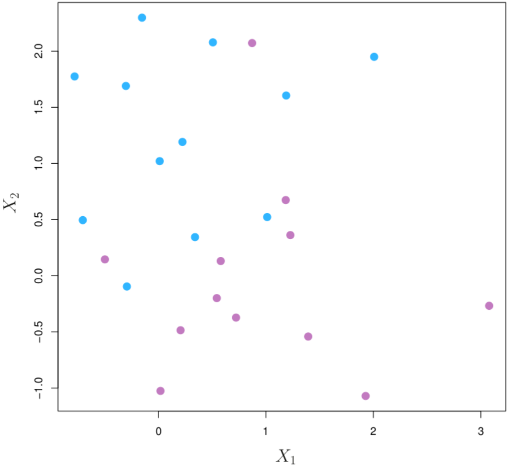
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M \;\; \forall\, i = 1, \ldots, n$$

- M is making sure every point is at least distance M from the hyperplane
- Because of the y_i multiplied, this also requires that the point is on the correct side of the hyperplane
- Last eq forces points to be on the correct side of the hyperplane if $M > 0$
- Making M as big as possible is the maximal margin hyperplane

Section 2

Issues with Maximal Margin Classifier

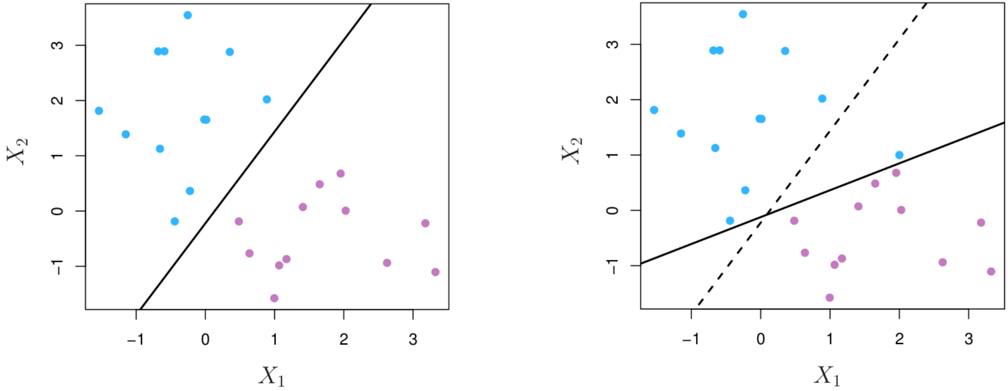
But what if....



- *...there is no separating hyperplane*
- *Means that there is no solution for $M > 0$*

Sensitivity to new points

Even if there is a spearating hyperplane might not be good



- Add a blue observation on the right
- Dramatic shift in maximal margin hyperplane