Ch 9.1: Maximum Margin Classifier

Lecture 25 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Nov 16, 2022

Announcements

Last time:

- Ch 8: Random Forests

This lecture:

- Maximal Margin Classifier
- No jupyter notebook for this class

Announcements:

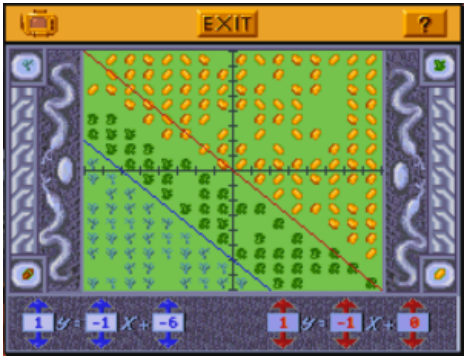
- HW #8 Due Friday
- No class (virtual OH only) Weds, Nov 23

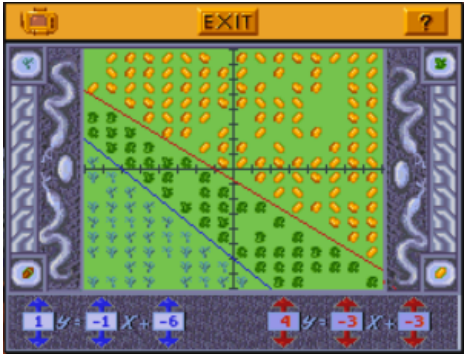
20	F	Nov 4	Polynomial & Step Functions.	7.1,7.2	
21	M	Nov 7	Step Functions	7.2	
22	W	Nov 9	Basis functions, Regression Splines	7.3,7.4	
23	F	Nov 11	Decision Trees	8.1	HW #7 Due
24	M	Nov 14	Random Forests	8.2.1, 8.2.2	
25	W	Nov 16	Maximal Margin Classifier	9.1	
26	F	Nov 18	SVC	9.2	HW #8 Due
27	M	Nov 21	SVM	9.3, 9.4, 9.5	
28	W	Nov 23	Extended virtual office hours		
	F	Nov 25	No class - Thanksgiving		
29	M	Nov 28	Single layer NN	10.1	HW #9 Due
30	W	Nov 30	Multi Layer NN	10.2	
31	F	Dec 2	CNN	10.3	
32	M	Dec 5	Unsupervised Learning & Clustering	12.1, 12.4	HW #10 Due
	W	Dec 7	Review		
	F	Dec 9	Midterm #3	Bring your cheat sheet and a non-internet-connected calculator	

Section 1

Maximal Margin Classifier

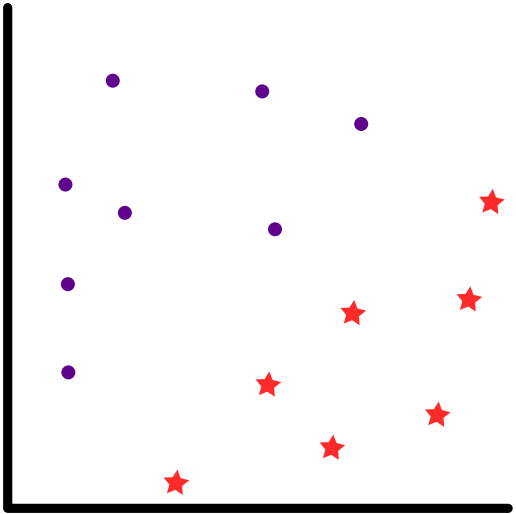
The Island of Dr Brain





https://classicreload.com/island_of_dr_brain.html

The goal

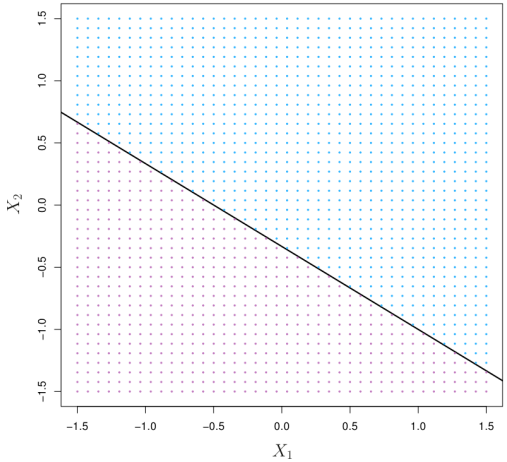


What is a hyperplane?

Mathematical definition of a hyperplane

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0$$

Hyperplane for $p = 2$

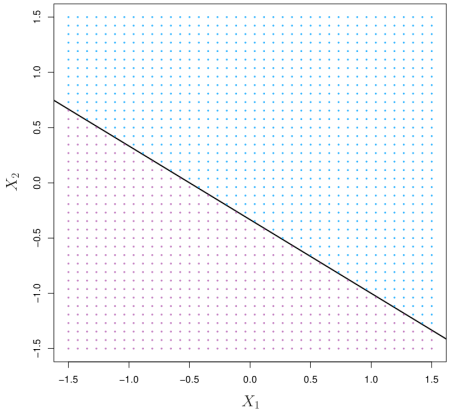


$$1 + 2X_1 + 3X_2 = 0$$

There are two sides to every hyperplane

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

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



Classification Setup

Data matrix:

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

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

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

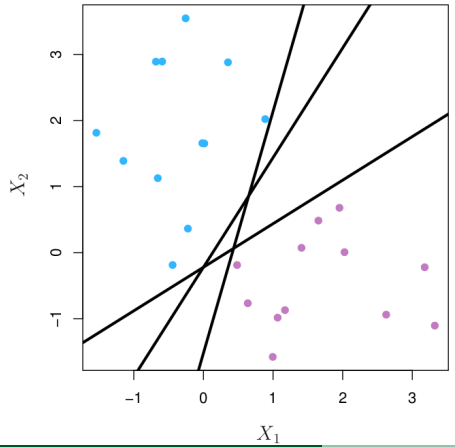
$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \cdots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

Separate out a test observation

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

Separating Hyperplane

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



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:

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

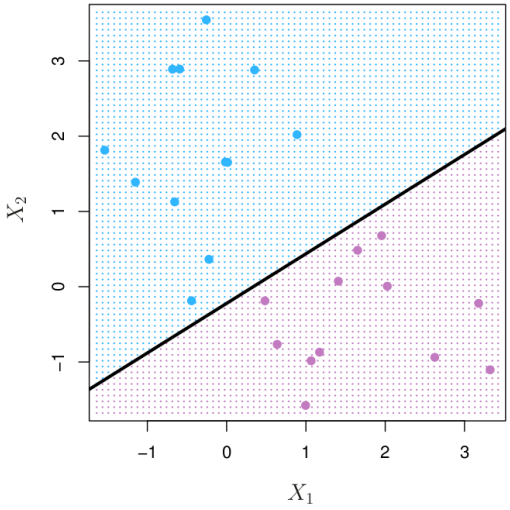
Separating hyperplane becomes a classifier

If you have a separating hyperplane:

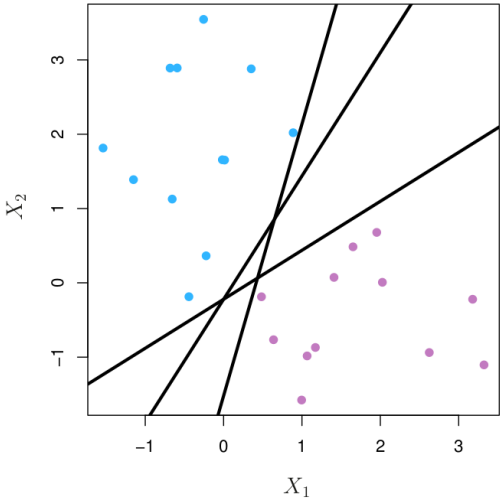
- Check

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

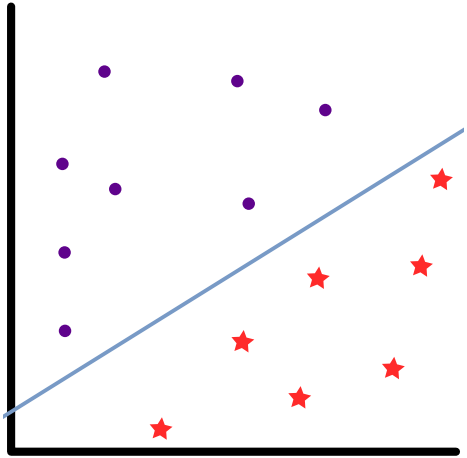
- If positive, assign $\hat{y} = 1$
- If negative, assign $\hat{y} = -1$

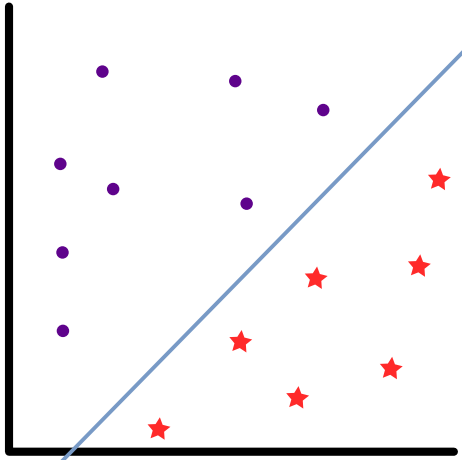


How do we pick?

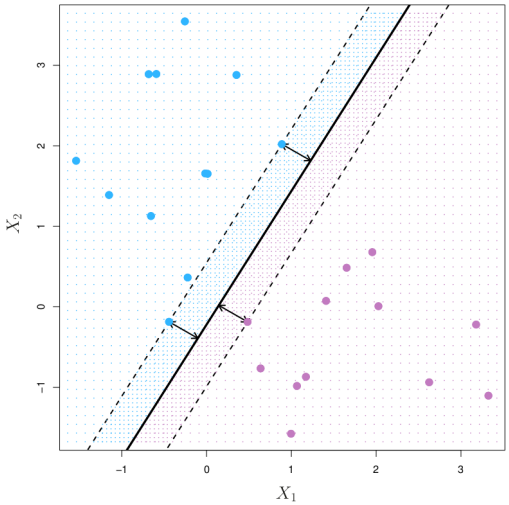


Distance from an observation to a hyperplane

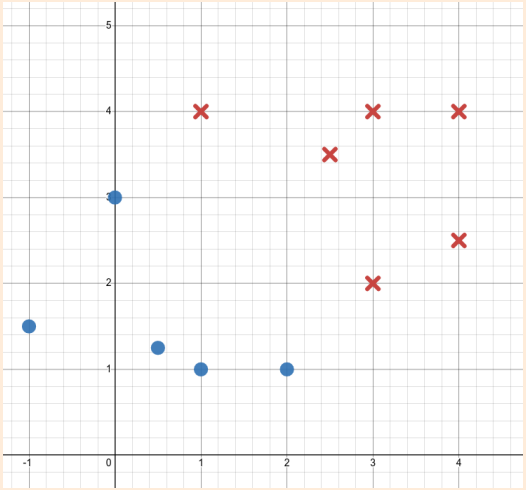




Maximal margin classifier



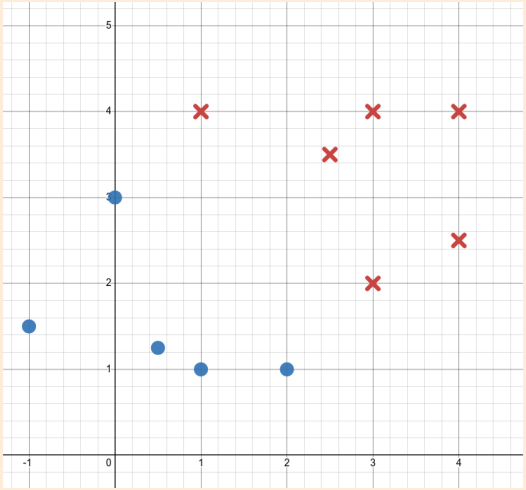
Example



- 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?

desmos.com/calculator/lqms253gfq

Extra work space



desmos.com/calculator/lqms253gfq

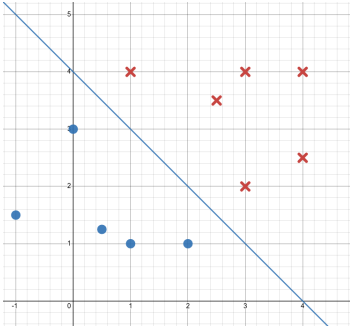
Mathematical Formulation

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

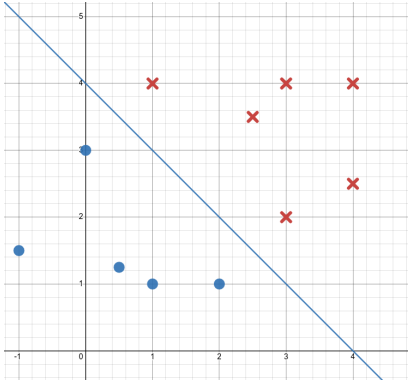
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

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

First constraint



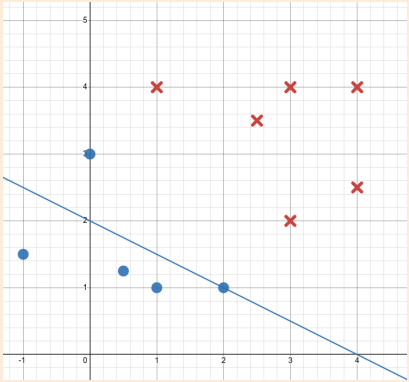
Second constraint



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

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

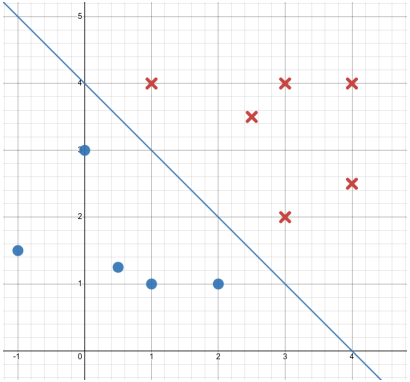
An example with a bad choice of hyperplane



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

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

Second constraint extra space



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

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

Mathematical Formulation

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

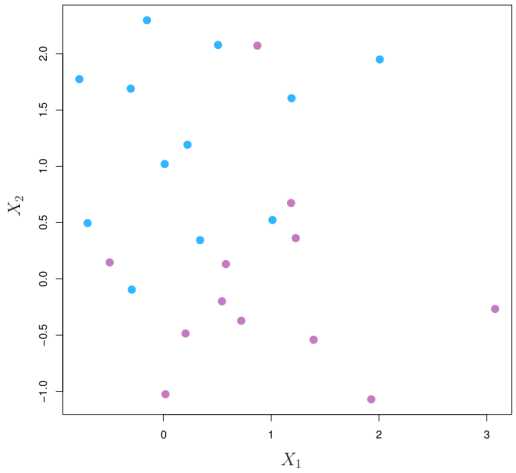
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

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

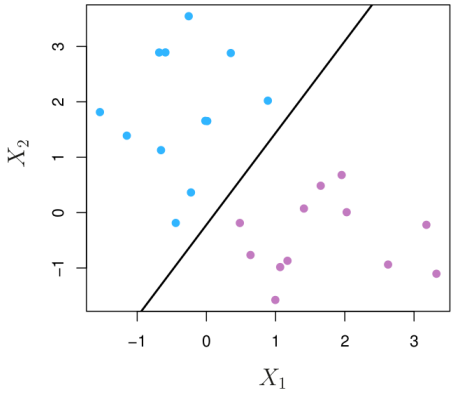
Section 2

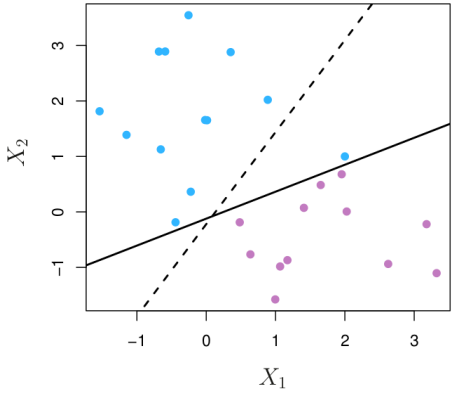
Issues with Maximal Margin Classifier

But what if....



Sensitivity to new points





Next time

20	F	Nov 4	Polynomial & Step Functions.	7.1,7.2	
21	M	Nov 7	Step Functions	7.2	
22	W	Nov 9	Basis functions, Regression Splines	7.3,7.4	
23	F	Nov 11	Decision Trees	8.1	HW #7 Due
24	M	Nov 14	Random Forests	8.2.1, 8.2.2	
25	W	Nov 16	Maximal Margin Classifier	9.1	
26	F	Nov 18	SVC	9.2	HW #8 Due
27	M	Nov 21	SVM	9.3, 9.4, 9.5	
28	W	Nov 23	Extended virtual office hours		
	F	Nov 25	No class - Thanksgiving		
29	M	Nov 28	Single layer NN	10.1	HW #9 Due
30	W	Nov 30	Multi Layer NN	10.2	
31	F	Dec 2	CNN	10.3	
32	M	Dec 5	Unsupervised Learning & Clustering	12.1, 12.4	HW #10 Due
	W	Dec 7	Review		
	F	Dec 9	Midterm #3	Bring your cheat sheet and a non-internet-connected calculator	