

Ch 9.1: Maximum Margin Classifier

Lecture 26 - CMSE 381

Prof. Guanqun Cao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Wed, Nov 5, 2025

Announcements

Last time:

- Ch 8: Random Forests

This lecture:

- Maximal Margin Classifier
- No jupyter notebook for this class

Announcements:

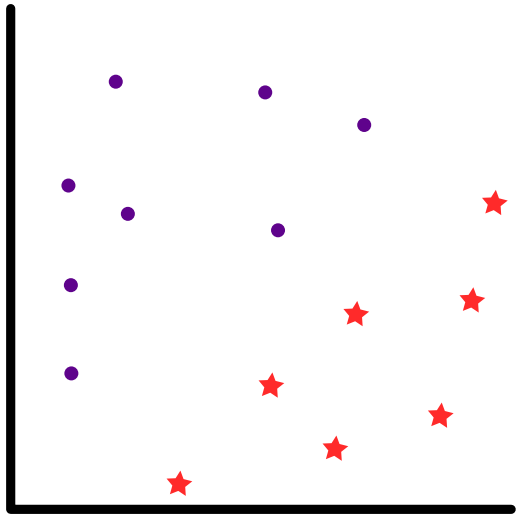
- HW #8 Due Sunday 11/9

	F	10/17	Review		
	M	10/20	Fall Break		
	W	10/22	Midterm #2		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2
25	M	11/3	Random Forests	8.2.1, 8.2.2	
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16
31	M	11/17	CNN	10.3	
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	Virtual: Project Office Hours		HW #9 Due Sun 11/23
	M	11/24	Review		
	W	11/26	Midterm #3		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		
	F	12/5			Project Due

Section 1

Maximal Margin Classifier

The goal

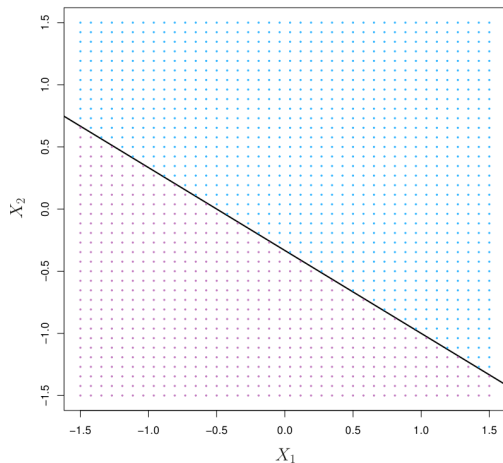


What is a hyperplane?

Mathematical definition of a hyperplane

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

Hyperplane for $p = 2$

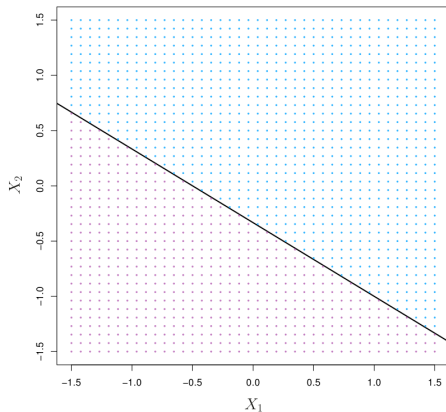


$$1 + 2X_1 + 3X_2 = 0$$

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:

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

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

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

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

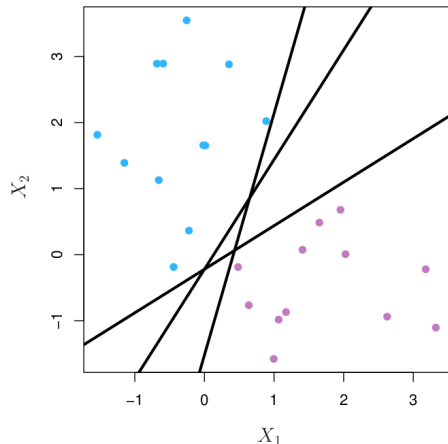
Separate out a test observation

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

Separating Hyperplane

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

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



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 :

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

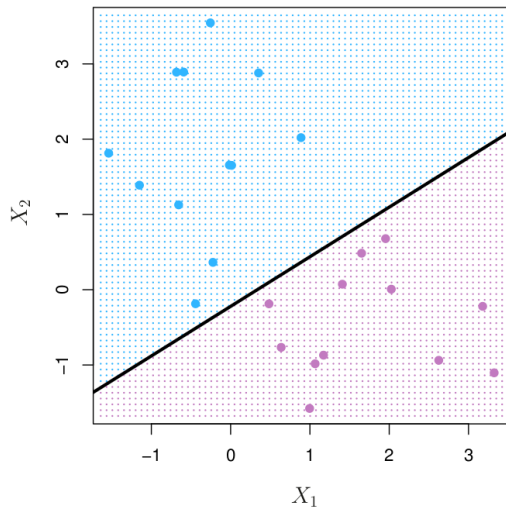
Separating hyperplane becomes a classifier

If you have a separating hyperplane:

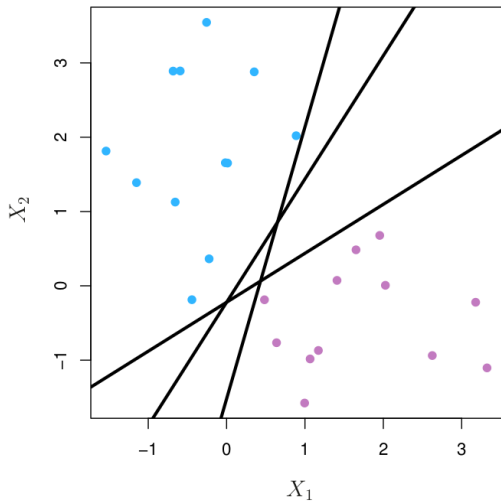
- Check

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

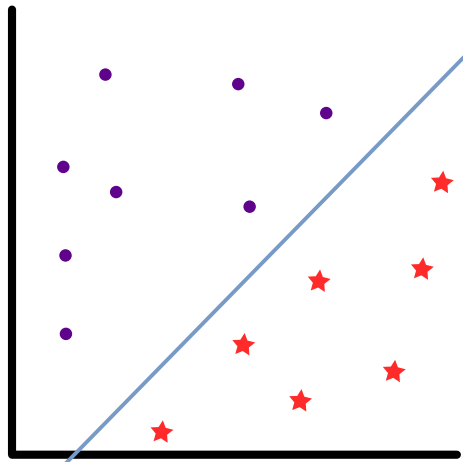
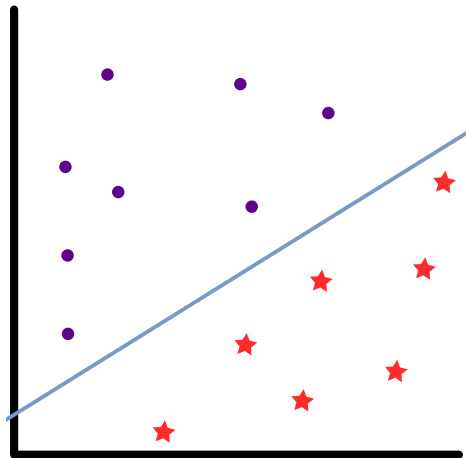
- If positive, assign $\hat{y} = 1$
- If negative, assign $\hat{y} = -1$



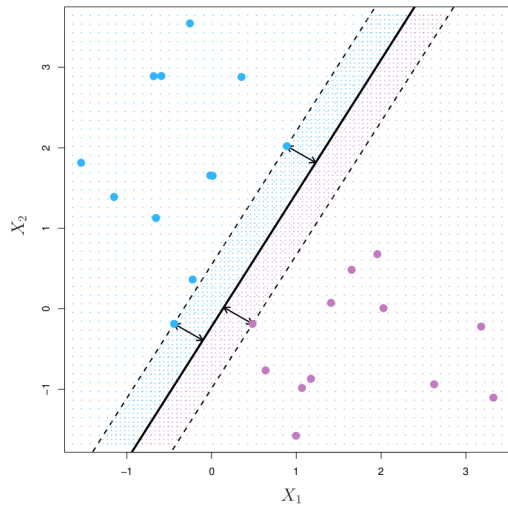
How do we pick?



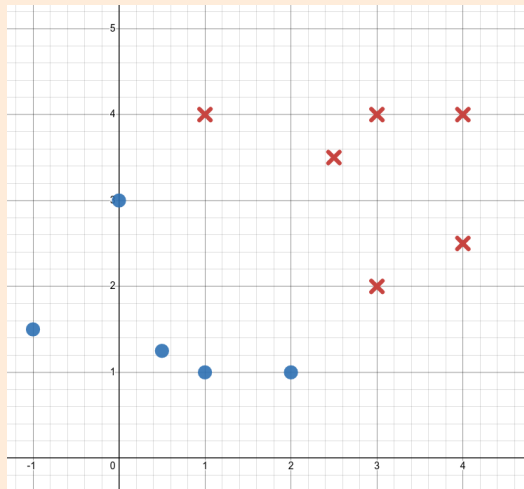
Distance from an observation to a hyperplane



Maximal margin classifier



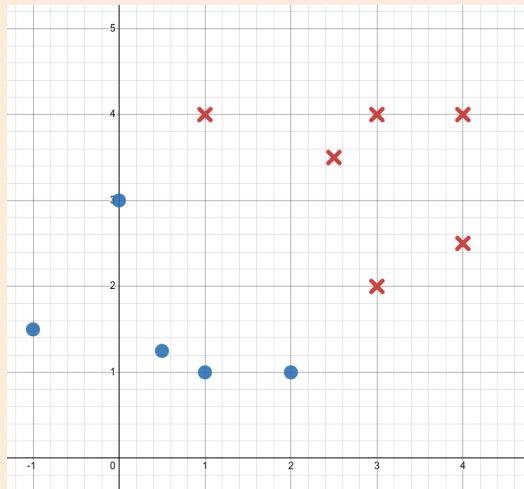
Example



- 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?

<https://www.desmos.com/calculator/tklbommiwz>

Extra work space



Respond to PollEv question.

<https://www.desmos.com/calculator/tklbommiwz>

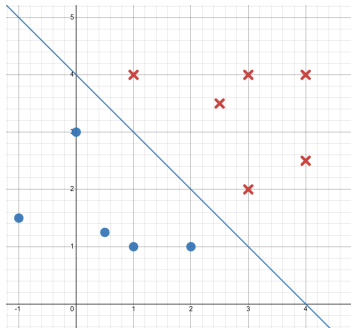
Mathematical Formulation

$$\underset{\beta_0, \beta_1, \dots, \beta_p, M}{\text{maximize}} \quad M$$

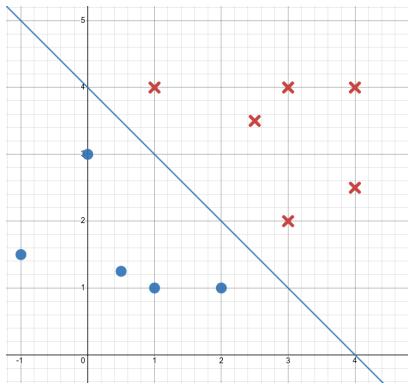
$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1,$$

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

First constraint



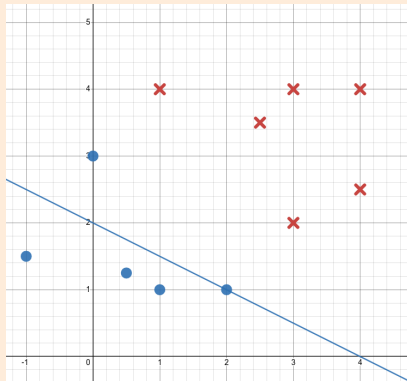
Second constraint



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

- Blue circles: $y_i = -1$
- Red Xs: $y_i = 1$
- $-2\sqrt{2} + \frac{\sqrt{2}}{2}X_1 + \frac{\sqrt{2}}{2}X_2 = 0$

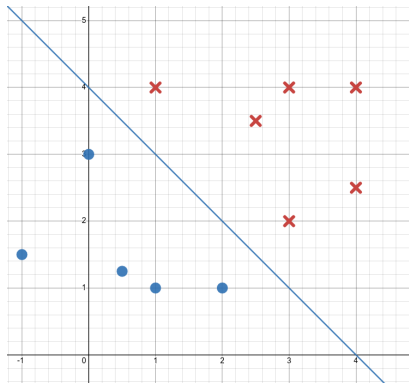
An example with a bad choice of hyperplane



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: $y_i = 1$
- $-\frac{4}{\sqrt{5}} + \frac{1}{\sqrt{5}}X_1 + \frac{2}{\sqrt{5}}X_2 = 0$

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: $y_i = 1$
- $-2\sqrt{2} + \frac{\sqrt{2}}{2}x_1 + \frac{\sqrt{2}}{2}x_2 = 0$

Mathematical Formulation

$$\underset{\beta_0, \beta_1, \dots, \beta_p, M}{\text{maximize}} \quad M$$

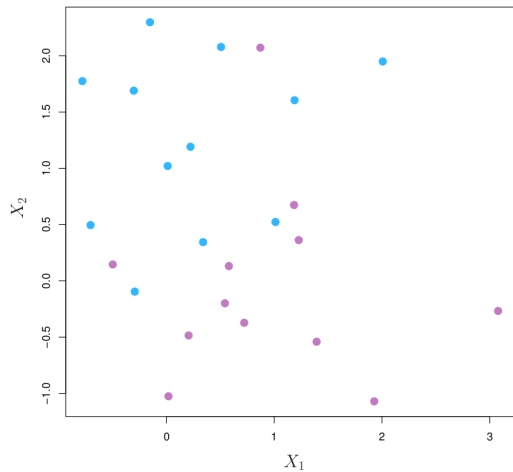
$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1,$$

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

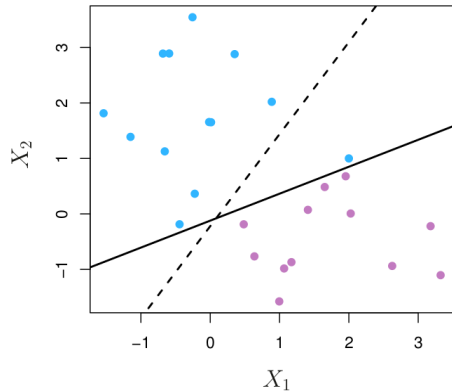
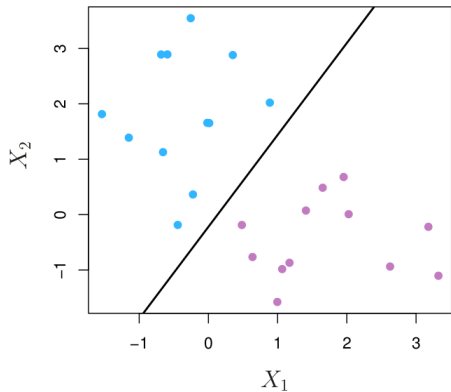
Section 2

Issues with Maximal Margin Classifier

But what if....



Sensitivity to new points



Next time

	F	10/17	Review		
	M	10/20	Fall Break		
	W	10/22	Midterm #2		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2
25	M	11/3	Random Forests	8.2.1, 8.2.2	
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16
31	M	11/17	CNN	10.3	
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	Virtual: Project Office Hours		HW #9 Due Sun 11/23
	M	11/24	Review		
	W	11/26	Midterm #3		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		
	F	12/5			Project Due

Q of the Day:

What is the relationship between “support vectors” and “margin”?