Ch 9.1: Maximum Margin Classifier

Lecture 27 - CMSE 381

Prof. Elizabeth Munch

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

Mon, Nov 13, 2023

Announcements

Last time:

- Ch 8: Random Forests

This lecture:

- Maximal Margin Classifier
- No jupyter notebook for this class

Announcements:

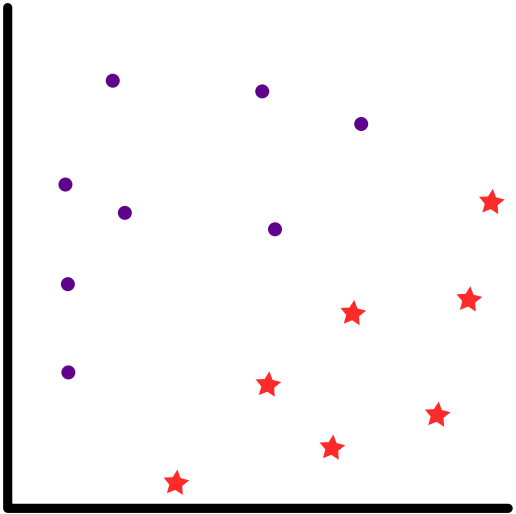
- HW #7....
- No class (virtual OH only) Weds, Nov 22

Status	Lec #	Date			Reading	Homeworks
		Mon	Oct 23	No class - Fall break		
		Wed	Oct 25	**Midterm #2**		
Done	20	Fri	Oct 27	Dimension Reduction	6.3	
Done	21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
Done	22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
Pushed	23	Fri	Nov 3	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
	24	Mon	Nov 6	Regression Splines	7.4	~~HW #6 Due~~
	25	Wed	Nov 8	Decision Trees	8.1	HW #6 Due
	26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
	27	Mon	Nov 13	Maximal Margin Classifier	9.1	
	28	Wed	Nov 15	SVC	9.2	
	29	Fri	Nov 17	SVM	9.3, 9.4	
	30	Mon	Nov 20	Single layer NN	10.1	
	31	Wed	Nov 22	Virtual: Project office hours		
		Fri	Nov 24	No class - Thanksgiving		
		Mon	Nov 27	**Review**		

Section 1

Maximal Margin Classifier

The goal

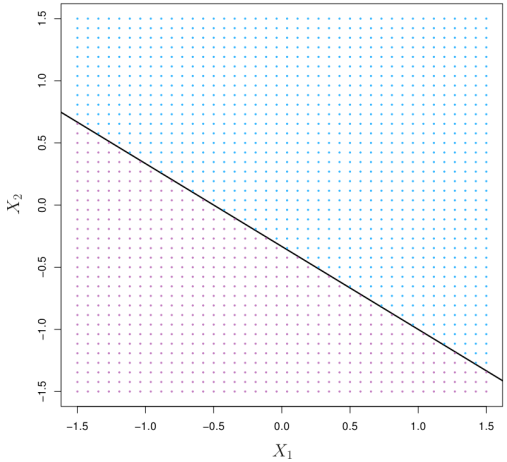


What is a hyperplane?

Mathematical definition of a hyperplane

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

Hyperplane for $p = 2$

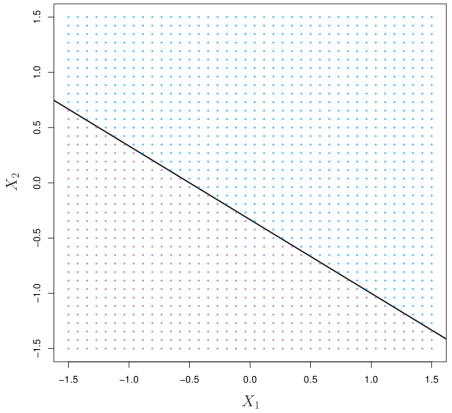


$$1 + 2X_1 + 3X_2 = 0$$

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:

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

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

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

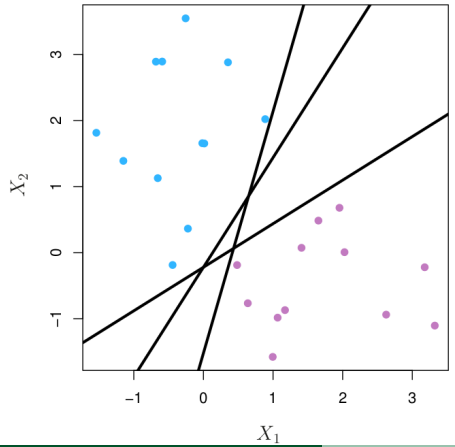
$$x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix}, \cdots, x_n = \begin{pmatrix} x_{n1} \\ \vdots \\ x_{np} \end{pmatrix}$$

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:

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

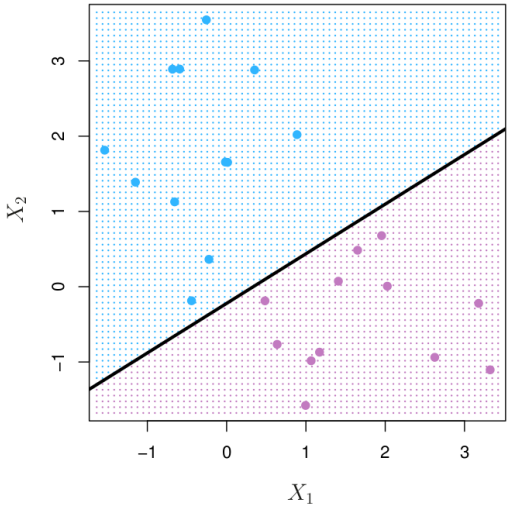
Separating hyperplane becomes a classifier

If you have a separating hyperplane:

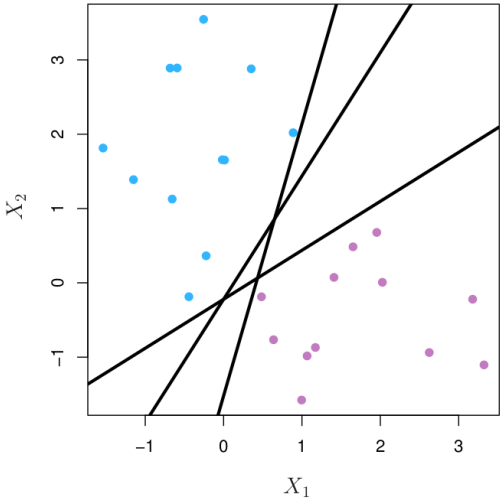
- Check

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

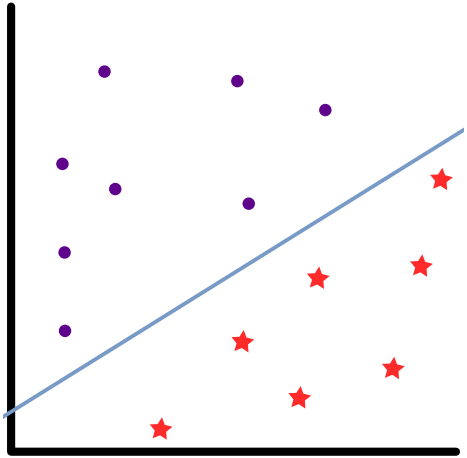
- If positive, assign $\hat{y} = 1$
- If negative, assign $\hat{y} = -1$

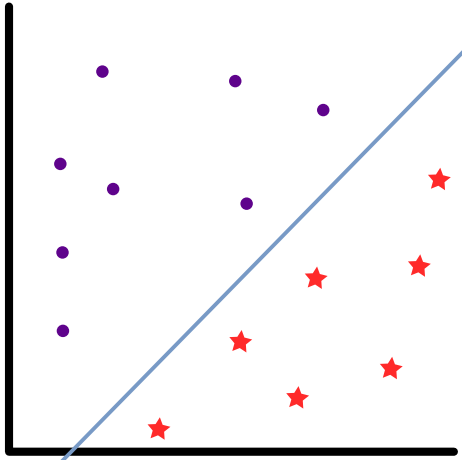


How do we pick?

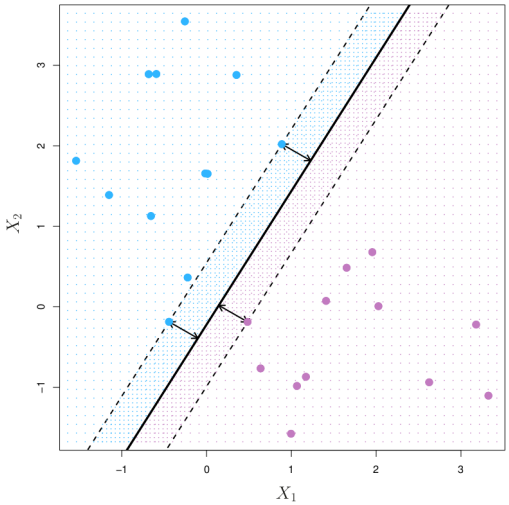


Distance from an observation to a hyperplane

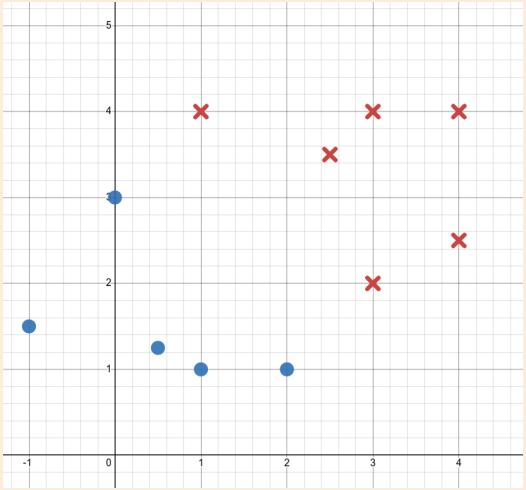




Maximal margin classifier



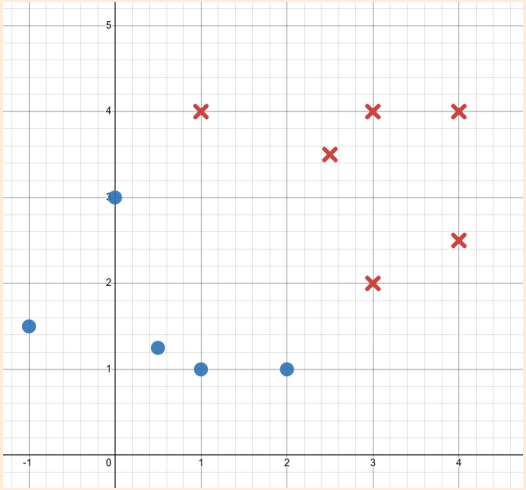
Example



- 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

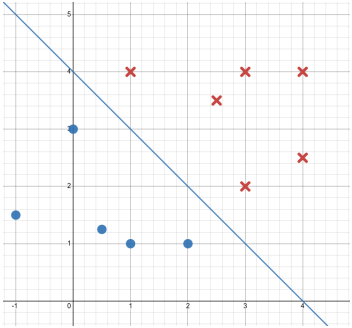
Mathematical Formulation

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

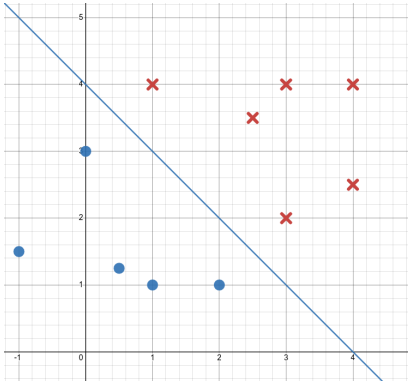
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

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

First constraint



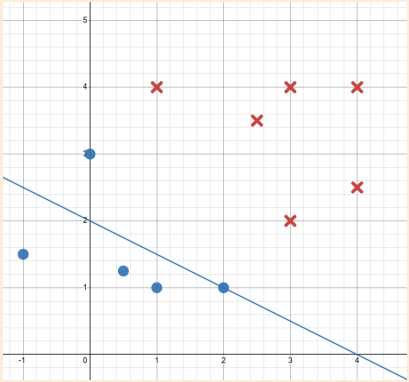
Second constraint



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

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

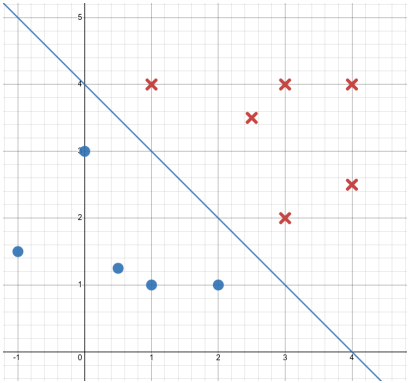
An example with a bad choice of hyperplane



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

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

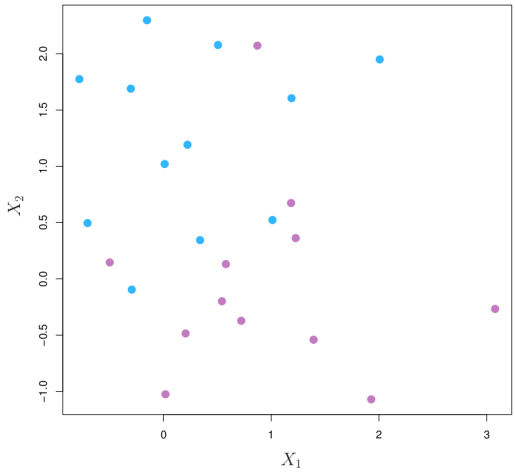
$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

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

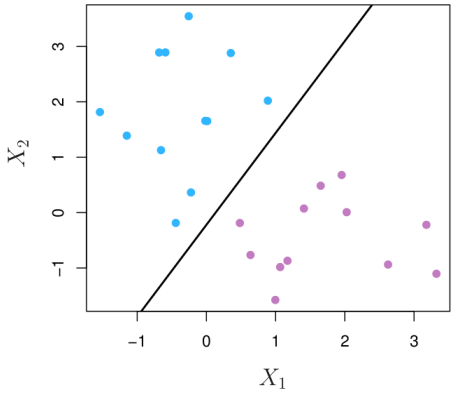
Section 2

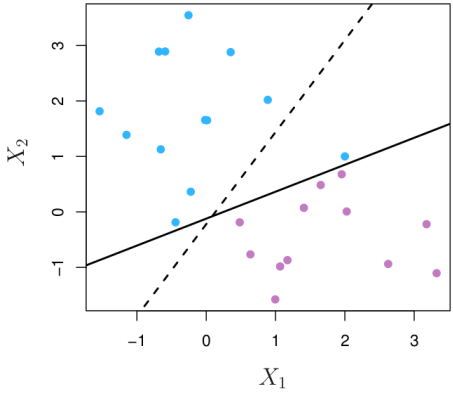
Issues with Maximal Margin Classifier

But what if....



Sensitivity to new points





Next time

Status	Lec #	Date			Reading	Homeworks
		Mon	Oct 23	No class - Fall break		
		Wed	Oct 25	**Midterm #2**		
Done	20	Fri	Oct 27	Dimension Reduction	6.3	
Done	21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
Done	22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
Pushed	23	Fri	Nov 3	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
	24	Mon	Nov 6	Regression Splines	7.4	~~HW #6 Due~~
	25	Wed	Nov 8	Decision Trees	8.1	HW #6 Due
	26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
	27	Mon	Nov 13	Maximal Margin Classifier	9.1	
	28	Wed	Nov 15	SVC	9.2	
	29	Fri	Nov 17	SVM	9.3, 9.4	
	30	Mon	Nov 20	Single layer NN	10.1	
	31	Wed	Nov 22	Virtual: Project office hours		
		Fri	Nov 24	No class - Thanksgiving		
		Mon	Nov 27	**Review**		
		Wed	Nov 29	**Midterm #3**		
	32	Fri	Dec 1	Multi Layer NN	10.2	
	33	Mon	Dec 4	CNN	10.3	
	34	Wed	Dec 6	Unsupervised Learning & Clustering	12.1, 12.4	
	35	Fri	Dec 8	Virtual: Project office hours		Project due