

Ch 9.3-4: Support Vector Machine

Lecture 29 - CMSE 381

Prof. Elizabeth Munch. Substitute: Elena Wang

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Fri, Nov 15, 2024

Announcements

Last time:

- 9.2 Support Vector Classifier

This lecture:

- 9.3 Support Vector Machine

Announcements:

- HW #7 due Monday

Lec #	Date		Reading	HW
21	Mon 10/28	Polynomial & Step Functions	7.1,7.2	
22	Wed 10/30	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
23	Fri 11/1	Regression Splines	7.4	HW #6 Due Sun 11/3
24	Mon 11/4	Decision Trees	8.1	
25	Wed 11/6	Class Cancelled (Dr Munch out of town)		
26	Fri 11/8	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 11/10
27	Mon 11/11	Maximal Margin Classifier	9.1	
28	Wed 11/13	SVC	9.2	
29	Fri 11/15	SVM	9.3, 9.4	HW #8 Due Sun 11/17
30	Mon 11/18	Single layer NN	10.1	
31	Wed 11/20	Multi Layer NN	10.2	
32	Fri 11/22	CNN	10.3	HW #11 Due Sun 11/24
33	Mon 11/25	TBD: Unsupervised learning/clustering	12.1, 12.4?	
	Wed 11/27	Virtual: Project office hours		
	Fri 11/29	No class - Thanksgiving		
	Mon 12/2	Review		
	Wed 12/4	Midterm #3		
	Fri 12/6	No class - EGR Design Day		Project due

Section 1

Last Time

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

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

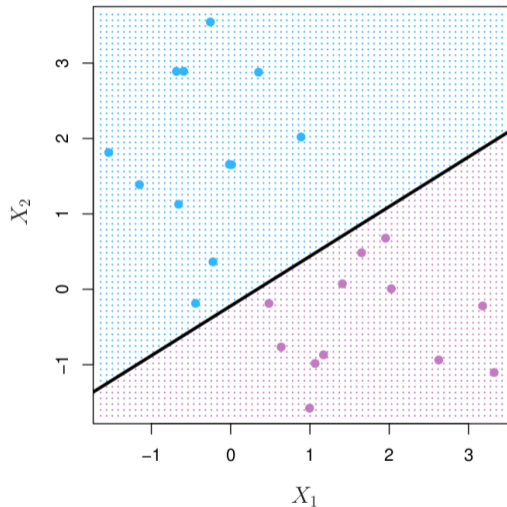
Hyperplane becomes a classifier

If you have a separating hyperplane:

- Check

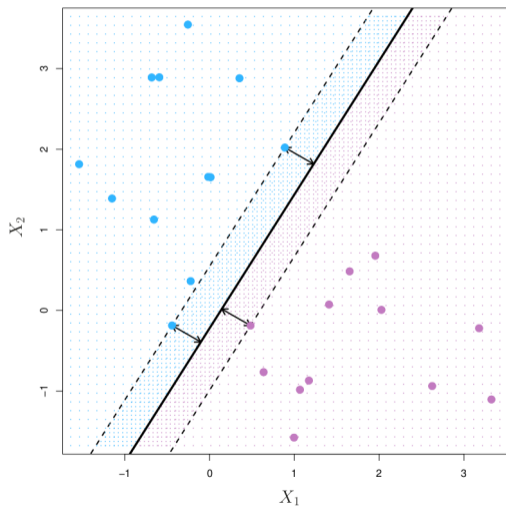
$$f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$$

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



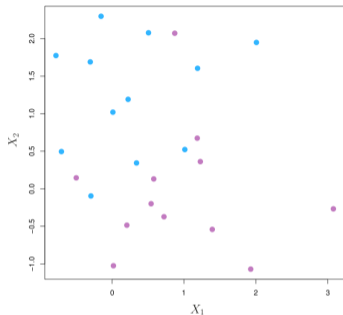
How do we pick? Old version

Maximal margin classifier

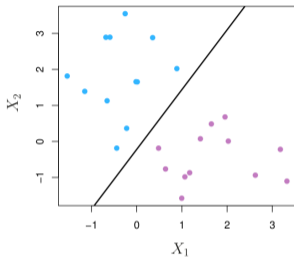


- For a hyperplane, the *margin* is the smallest distance from any data point to the hyperplane.
- Observations that are closest are called *support vectors*.
- The *maximal margin hyperplane* is the hyperplane with the largest margin
- The classifier built from this hyperplane is the *maximal margin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

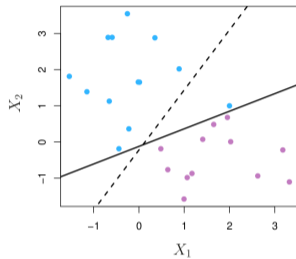
Issues



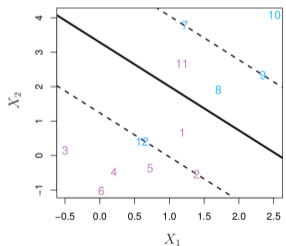
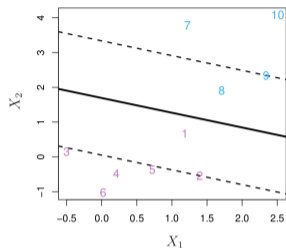
No separating hyperplane exists



Choice of hyperplane is sensitive to new points



Support Vector Classifier



$$\text{maximize } M$$

$$\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M$$

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

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

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

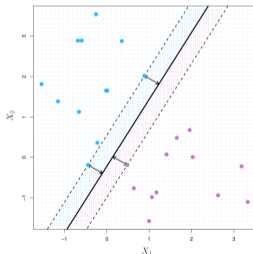
Two formulations side by side

Maximal Margin Classifier

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

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

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



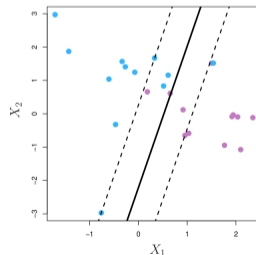
Support Vector Classifier

$$\text{maximize}_{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M} M$$

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

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

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



So many variables

$$\text{maximize}_{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n, M} M$$

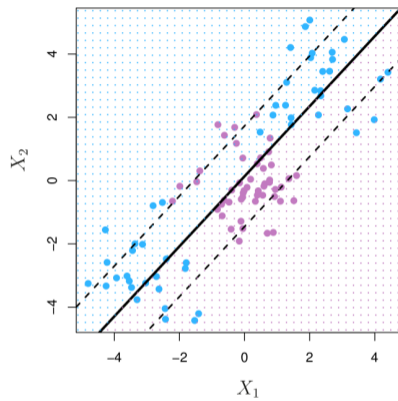
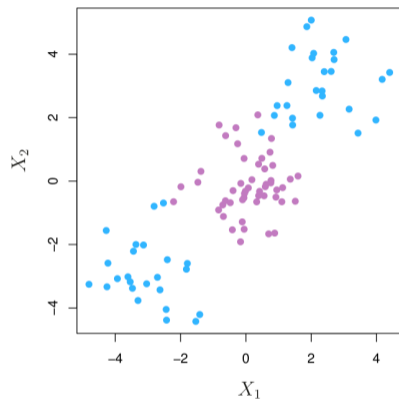
$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1,$$

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

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

- C is nonnegative tuning parameter
- M is the width of the margin
- $\epsilon_1, \dots, \epsilon_n$ are slack variables allowing observations to go to the other side

Limiting factor of SVC



Section 2

Support Vector Machine

Example of using more features

Want $2p$ features:

$$X_1, X_1^2, X_2, X_2^2, \dots, X_p, X_p^2$$

Optimization becomes:

$$\begin{aligned} & \underset{\beta_0, \beta_{11}, \beta_{12}, \dots, \beta_{p1}, \beta_{p2}, \epsilon_1, \dots, \epsilon_n, M}{\text{maximize}} && M \\ \text{subject to } & y_i \left(\beta_0 + \sum_{j=1}^p \beta_{j1} x_{ij} + \sum_{j=1}^p \beta_{j2} x_{ij}^2 \right) \geq M(1 - \epsilon_i), \\ & \sum_{i=1}^n \epsilon_i \leq C, \quad \epsilon_i \geq 0, \quad \sum_{j=1}^p \sum_{k=1}^2 \beta_{jk}^2 = 1. \end{aligned}$$

$$\langle a, b \rangle = \sum_{i=1}^r a_i b_i$$

Quick computations

What are the following?

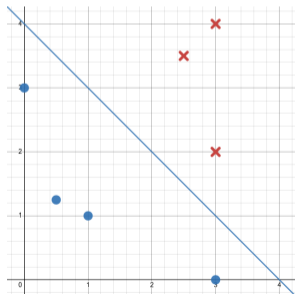
- $\langle (1, 1), (0, 3) \rangle$
- $\langle (1, 1), (3, 2) \rangle$
- $\langle (2, 3), (0, 3) \rangle$
- $\langle (2, 3), (3, 2) \rangle$

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

Example

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

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



What are the α_i 's?

Data point	α_i
(3, 4)	
(2.5, 3.5)	
(3, 2)	
(3, 0)	
(0, 3)	
(1, 1)	
(0.5, 1.25)	

What α_i 's are needed to write the hyperplane

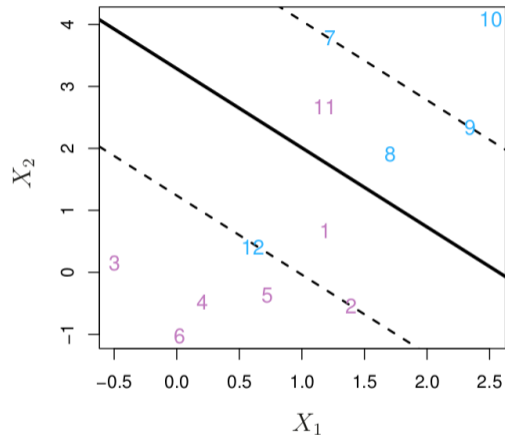
$$-2\sqrt{2} + \frac{\sqrt{2}}{18} \langle (X_1, X_2), (0, 3) \rangle + \frac{\sqrt{2}}{6} \langle (X_1, X_2), (3, 2) \rangle$$

of the previous page in the form

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle?$$

SVC via inner products of support vectors

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i \langle x, x_i \rangle$$

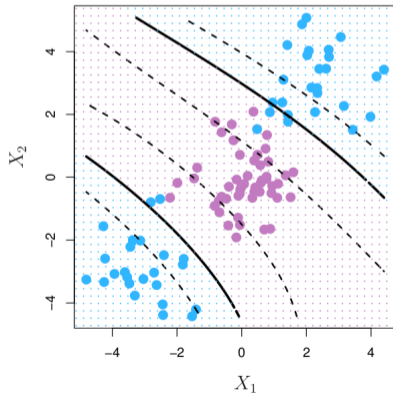


$$K(x_i, x'_j)$$

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i K(x, x_i)$$

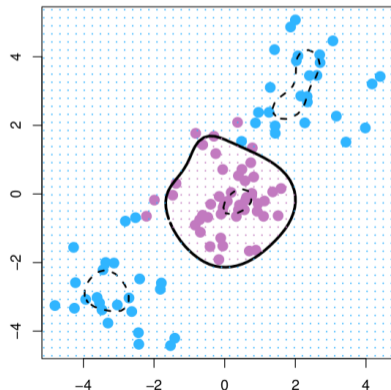
A polynomial kernel

$$K(x_i, x_{i'}) = \left(1 + \sum_{j=1}^p x_{ij}x_{i'j} \right)^d$$



A radial kernel

$$K(x_i, x_i') = \exp \left(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right)$$



Support Vector Machine

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i K(x, x_i)$$

Coding

Section 3

SVM with more than two classes

One-Vs-One Classification

Also called all-pairs

One-Vs-All Classification

$$f(x) = \beta_0 + \sum_{i \in \mathcal{S}} \alpha_i K(x, x_i)$$

Kernels

- Linear

$$K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} x_{i'j}$$

- Polynomial

$$K(x_i, x_{i'}) = \left(1 + \sum_{j=1}^p x_{ij} x_{i'j} \right)^d$$

- Radial

$$K(x_i, x_{i'}) = \exp \left(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right)$$

Next time

Lec #	Date		Reading	HW
21	Mon 10/28	Polynomial & Step Functions	7.1,7.2	
22	Wed 10/30	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
23	Fri 11/1	Regression Splines	7.4	HW #6 Due
24	Mon 11/4	Decision Trees	8.1	Sun 11/3
25	Wed 11/6	Class Cancelled (Dr Munch out of town)		
26	Fri 11/8	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 11/10
27	Mon 11/11	Maximal Margin Classifier	9.1	
28	Wed 11/13	SVC	9.2	
29	Fri 11/15	SVM	9.3, 9.4	HW #8 Due
30	Mon 11/18	Single layer NN	10.1	Sun 11/17
31	Wed 11/20	Multi Layer NN	10.2	
32	Fri 11/22	CNN	10.3	HW #11 Due Sun 11/24
33	Mon 11/25	TBD: Unsupervised learning/clustering	12.1, 12.4?	
	Wed 11/27	Virtual: Project office hours		
	Fri 11/29	No class - Thanksgiving		
	Mon 12/2	Review		
	Wed 12/4	Midterm #3		
	Fri 12/6	No class - EGR Design Day		Project due