

Ch 9.3-4: Support Vector Machine

Lecture 29 - CMSE 381

Prof. Mengsen Zhang

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Fri, April 4, 2025

Announcements

Last time:

- 9.2 Support Vector Classifier

This lecture:

- 9.3 Support Vector Machine

Announcements:

- HW #8 due Sunday 4/6

	M	3/17	Midterm #2		Sun 3/16
21	W	3/19	Polynomial & Step Functions	7.1-7.2	
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	M	3/24	Regression Splines	7.4	
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30
26	M	3/31	Maximal Margin Classifier	9.1	
27	W	4/2	SVC	9.2	
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6
29	M	4/7	Single Layer NN	10.1	
30	W	4/9	Multi Layer NN	10.2	
31	F	4/11	CNN	10.3	HW #9 Due Sun 4/13
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4	
33	W	4/16	Virtual Project Office Hours		
	F	4/18	Review		
	M	4/21	Midterm #3		
	W	4/23			
	F	4/25			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

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

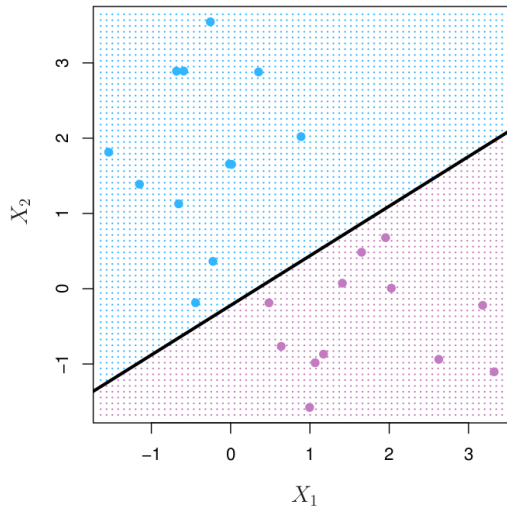
Hyperplane becomes a classifier

If you have a separating hyperplane:

- Check

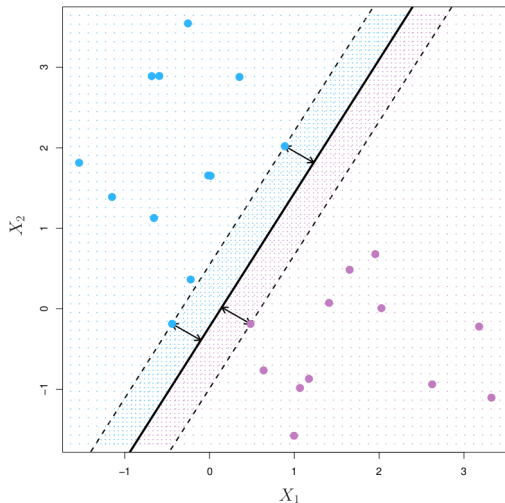
$$f(\mathbf{x}^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^*$$

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



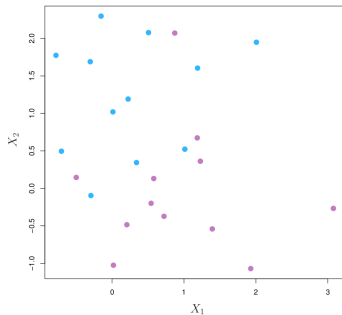
How do we pick? Old version

Maximal margin classifier

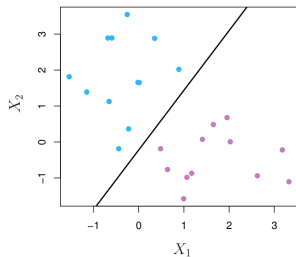


- For a hyperplane, the *margin* is the smallest distance from any data point to the hyperplane.
- Observations that are closest are called *support vectors*.
- The *maximal margin hyperplane* is the hyperplane with the largest margin
- The classifier built from this hyperplane is the *maximal margin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

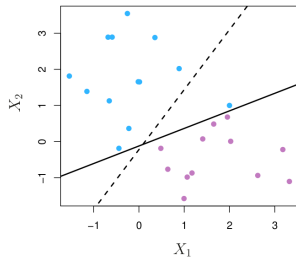
Issues



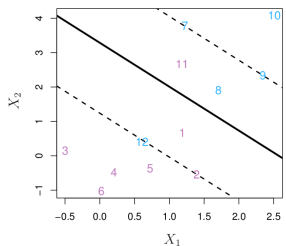
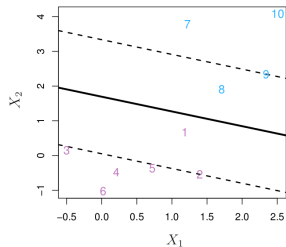
No separating hyperplane
exists



Choice of hyperplane is sensitive to new points



Support Vector Classifier



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

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

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

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

Test your understanding in PolIEv!

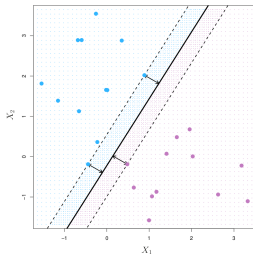
Two formulations side by side

Maximal Margin Classifier

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

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

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



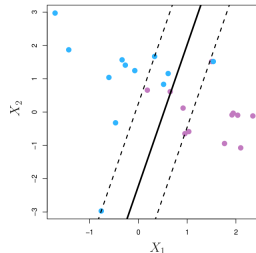
Support Vector Classifier

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

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

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

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



So many variables

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

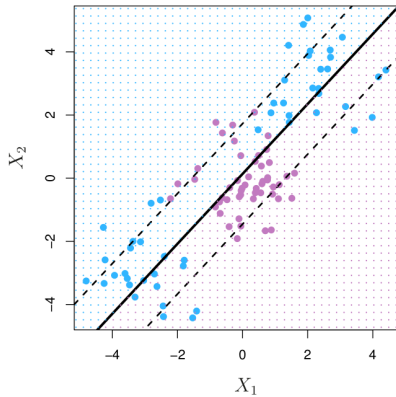
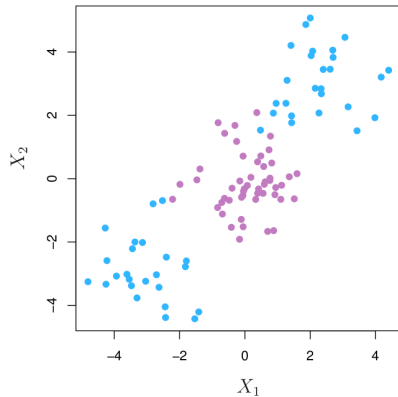
$$\text{subject to} \quad \sum_{j=1}^p \beta_j^2 = 1,$$

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

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

- 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:

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

$$y_i f(x_i) \geq M(1 - \epsilon_i)$$

$f(x_i)$ should give you some more general way of talking about how far is point x_i away from the decision boundary.

Inner products

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

Quick computations

What are the following?

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

SVC via inner products

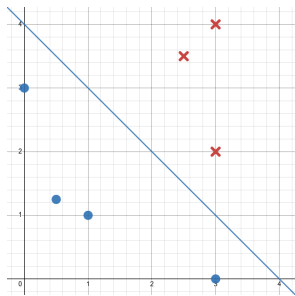
$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle$$

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

Example

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

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



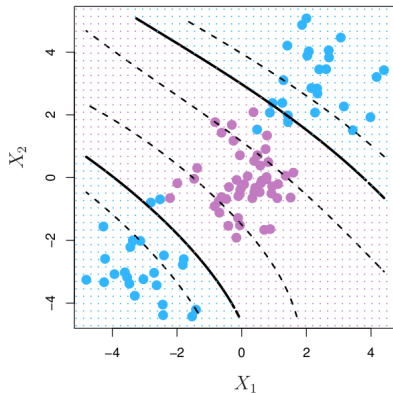
The kernel

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

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

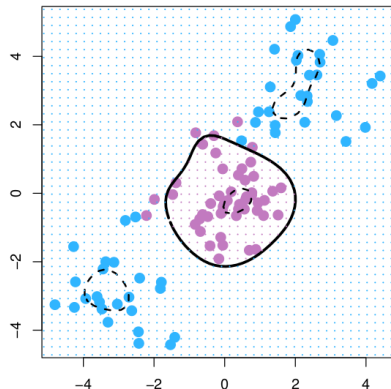
A polynomial kernel

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



A radial kernel

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



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

	M	3/17	Midterm #2		Sun 3/16
21	W	3/19	Polynomial & Step Functions	7.1-7.2	
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	M	3/24	Regression Splines	7.4	
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30
26	M	3/31	Maximal Margin Classifier	9.1	
27	W	4/2	SVC	9.2	
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6
29	M	4/7	Single Layer NN	10.1	
30	W	4/9	Multi Layer NN	10.2	
31	F	4/11	CNN	10.3	HW #9 Due Sun 4/13
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4	
33	W	4/16	Virtual: Project Office Hours		
	F	4/18	Review		
	M	4/21	Midterm #3		
	W	4/23			
	F	4/25			Project Due

Q of the day:

Why would you want to use SVM rather than SVC?