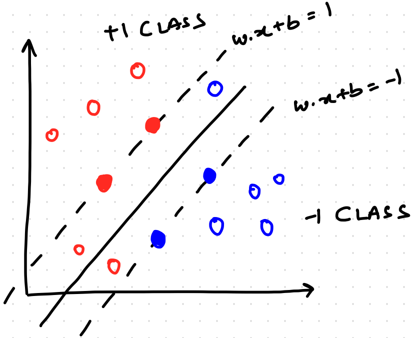
SVM Soft Margin Classification

Nipun Batra

IIT Gandhinagar

August 30, 2025

"SLIGHTLY" NON-SEPARABLE DATA



Pop Quiz #1

Answer this!

Why might we need a ”soft margin” SVM?

A) Data is perfectly linearly separable
B) Data has some noise and outliers
C) We want smaller margins
D) To avoid using kernels

Pop Quiz #1

Answer this!

Why might we need a "soft margin" SVM?

A) Data is perfectly linearly separable

B) Data has some noise and outliers

C) We want smaller margins

D) To avoid using kernels

Answer: B) Data has some noise and outliers - soft margin allows controlled violations.

Soft-Margin SVM

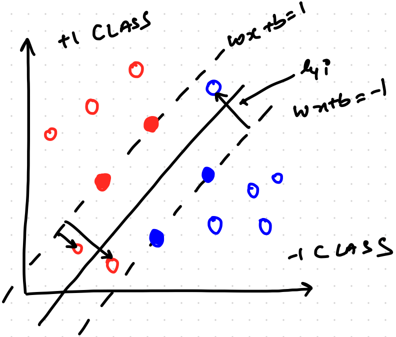
- Can we learn SVM for "slightly" non-separable data without projecting to a higher space?

Soft-Margin SVM

- Can we learn SVM for “slightly” non-separable data without projecting to a higher space?
- Introduce some “slack” (ξ_i) or loss or penalty for samples - allow some samples to be misclassified

"SLIGHTLY" NON-SEPARABLE DATA

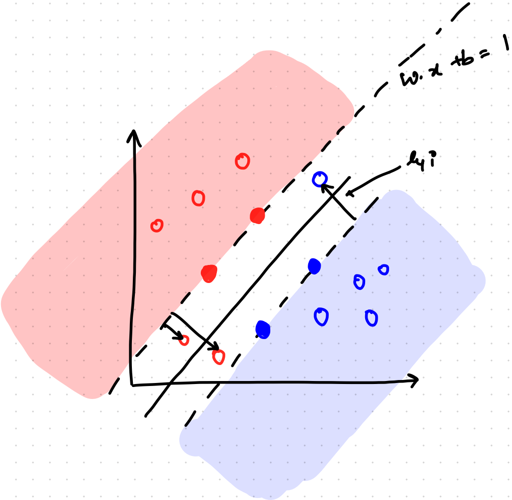
ξ_i: Distance from margin



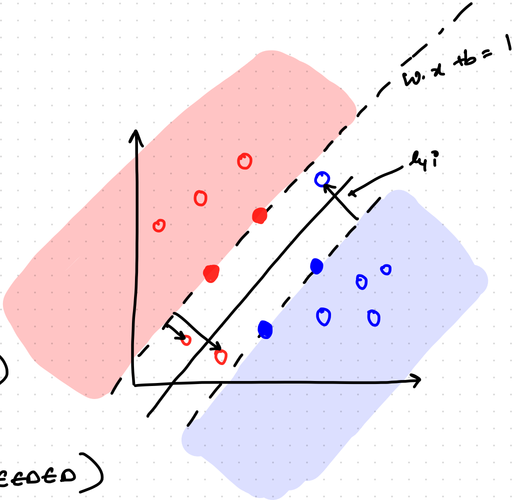
ZONE 1

$y_i = +1$

$w \cdot x_i + b \geq 1$



ZONE 1



$y_i(\vec{w} \cdot \vec{x_i} + b) > 1$

Loss$_i$ = 0 ($\mathcal{E}_i = 0$)

$\Downarrow$

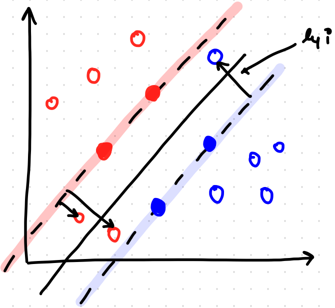
NO PENALTY
(OR SLACK NEEDED)

ZONE 2

$y_i(\vec{w}.\vec{x}_i + b) = 1$

$LOSS_i = 0$

$(\xi_i = 0)$



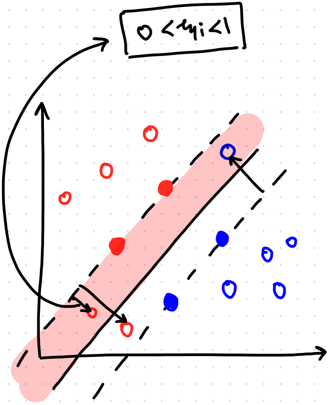
ZONE 3

$y_i(\vec{w}.\vec{x}_i + b) < 1$

$LOSS_i \neq 0$ $(0 < \xi_i < 1)$

POINT CORRECTLY CLASSIFIED

(BUT WRONG SIDE OF MARGIN)



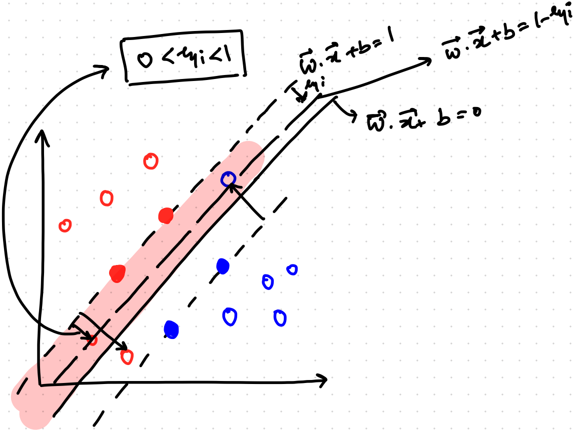
ZONE 3

$y_i(\vec{w}.\vec{x_i}+b) < 1$

$LOSS_i \neq 0$ $(0 < \xi_i < 1)$

POINT CORRECTLY CLASSIFIED

(BUT WRONG SIDE OF MARGIN)



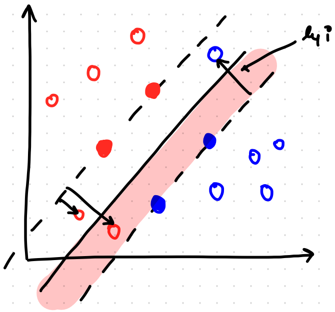
ZONE 4

$y_i(\vec{w}\cdot\vec{x_i}+b)<1$

POINT INCORRECTLY CLASSIFIED

$LOSS_i \neq 0$

$\xi_i > 1$



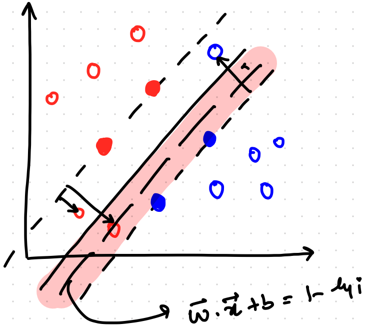
ZONE 4

$y_i(\vec{w}\cdot\vec{x_i}+b) < 1$

POINT INCORRECTLY CLASSIFIED

$LOSS_i \neq 0$

$\xi_i > 1$



Soft-Margin SVM

Change Objective

$\text{minimize}\, \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i$

s.t. $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$

Soft-Margin SVM

Change Objective

$\text{minimize}\ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$

s.t. $y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i$

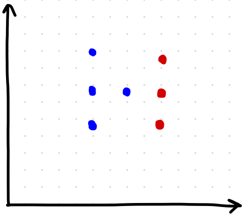
In Dual:

$$\text{minimize}\sum_{i=1}^{n}\boldsymbol{\alpha}_i - \sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j y_i y_j \mathbf{x}_i\cdot\mathbf{x}_j$$

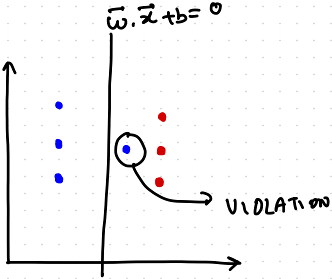
s.t.

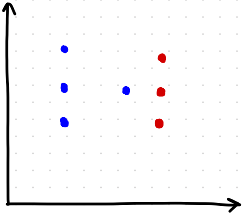
$$0 \leq \boldsymbol{\alpha}_i \leq C \quad \& \quad \sum_{i=1}^{n}\boldsymbol{\alpha}_i y_i = 0$$

BIAS-VARIANCE TRADE-OFF



BIAS-VARIANCE TRADE-OFF





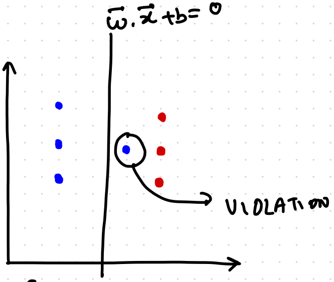
LOW C

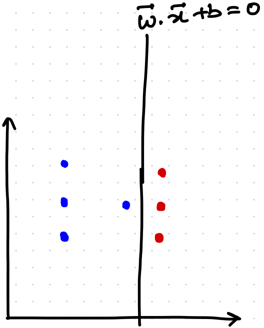
LOW PENALTY FOR VIOLATION

HIGH TRAIN ERROR

HIGH BIAS

BIAS-VARIANCE TRADE-OFF





LOW C

LOW PENALTY FOR VIOLATION

HIGH TRAIN ERROR

HIGH BIAS

BIG MARGIN

HIGH C

HIGH PENALTY FOR VIOLATION

HIGH VARIANCE

SMALL MARGIN

Pop Quiz #2

Answer this!

What happens when the regularization parameter C is very large?

A) The model becomes more tolerant to misclassifications
B) The model tries to classify all training points correctly
C) The margin becomes larger
D) Regularization increases

Pop Quiz #2

Answer this!

What happens when the regularization parameter C is very large?

A) The model becomes more tolerant to misclassifications

B) The model tries to classify all training points correctly

C) The margin becomes larger

D) Regularization increases

Answer: **B) The model tries to classify all training points correctly** - high variance!

Bias Variance Trade-off for Soft-Margin SVM

Low C $\implies$ Higher train error (higher bias)

High C $\implies$ Very sensitive to datasete (high variance)

Soft-Margin SVM

If C $\to 0$

Objective $\to \mathrm{minimize}\ \frac{1}{2}\|\mathbf{w}\|^2$

$\implies$ Choose large margin (without worrying for ξ_is)

$$\boxed{\text{Recall: Margin} = \frac{2}{\|\mathbf{w}\|}}$$

If C $\to \infty$ (or very large) Objective $\to \mathrm{minimize}\ C \sum \xi_i$ or choose $\mathbf{w}$, b, s.t. ξ_i is small!

Pop Quiz #3

Answer this!

What is the equivalent of hard margin?

A) $C \to 0$

B) $C \to \infty$

Pop Quiz #3

Answer this!

What is the equivalent of hard margin?

A) $\text{C} \to 0$

B) $\text{C} \to \infty$

Answer: B) C $\to \infty$ - No violations allowed!

Pop Quiz #4

Answer this!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

A) On the margin boundary

B) Correctly classified but within margin

C) Misclassified

D) Outside both margins

Pop Quiz #4

Answer this!

For a support vector with slack variable $\xi_i = 1.5$, this point is:

A) On the margin boundary

B) Correctly classified but within margin

C) Misclassified

D) Outside both margins

Answer: C) Misclassified - since $\xi_i > 1$!

Soft-Margin SVM

Types of support vectors:

- Zone 2: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
- Zone 3: $0 < \xi_i < 1$ (correctly classified)
- Zone 4: $\xi_i > 1$ (Misclassified)

$\therefore$ As C increases, $\#$ support vectors decreases

Notebook: SVM-soft-margin

SVM Formulation in the Loss + Penalty Form

Objective:

$$\text{minimize}\,\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

Now:

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 1 - y_i(\mathbf{w}\cdot\mathbf{x}_i + b)$$

But $\xi_i \geq 0$

$$\therefore \xi_i = \max\left[0, 1 - y_i(\mathbf{w}\cdot\mathbf{x}_i + b)\right]$$

Pop Quiz #5

Answer this!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ **is:**

A) Convex and differentiable everywhere

B) Convex but not differentiable at one point

C) Non-convex but differentiable

D) Neither convex nor differentiable

Pop Quiz #5

Answer this!

The hinge loss function $\max[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)]$ **is:**

A) Convex and differentiable everywhere

B) Convex but not differentiable at one point

C) Non-convex but differentiable

D) Neither convex nor differentiable

Answer: B) Convex but not differentiable at one point
- at $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$!

SVM Formulation in the Loss + Penalty Form

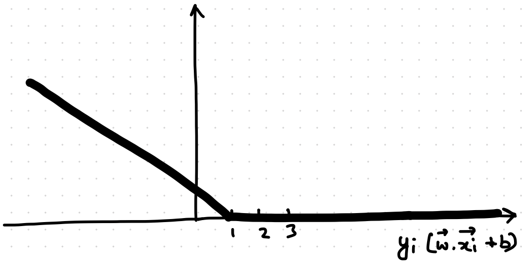
$\therefore$ Objective is:

$$\text{minimize } C \sum \xi_i + \frac{1}{2}\|\mathbf{w}\|^2$$

$$\implies \text{minimize } C \sum_{i=1}^{N} \max\left[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)\right] + \frac{1}{2}\|\mathbf{w}\|^2$$

$$\implies \text{minimize} \underbrace{\sum_{i=1}^{N} \max\left[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)\right]}_{\text{Loss}} + \underbrace{\frac{1}{2C}\|\mathbf{w}\|^2}_{\text{Regularisation}}$$

HINGE LOSS



Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max\left[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)\right]$

- Case I $\quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max\left[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)\right]$

- Case I $\quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$
- Case II
  $$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1$$
  $Loss_i = 0$

Loss Function for Sum (Hinge Loss)

Loss function is $\sum_{i=1}^{N} \max\left[0, 1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b)\right]$

- Case I $\quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) = 1$
  Lies on Margin: $Loss_i = 0$
- Case II
  $$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1$$
  $Loss_i = 0$
- Case III
  $$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) < 1$$
  $Loss_i \neq 0$

Hinge Loss Continued

Q) Is hinge loss convex and differentiable?

Convex: ✓

Differentiable: X

Subgradient: ✓

SVM Loss is Convex

Hinge Loss $\sum(\max[0, (1 - y_i(\mathbf{w} \cdot \mathbf{x}_i + b))]$ is convex

Penalty $\frac{1}{2}\|\mathbf{w}\|^2$ is convex

$\therefore$ SVM loss is convex