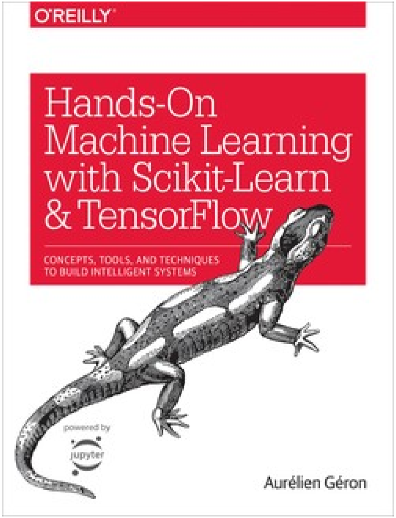
Machine Learning in Geophysics

Lecture 2 – Support Vector Machines

Resources



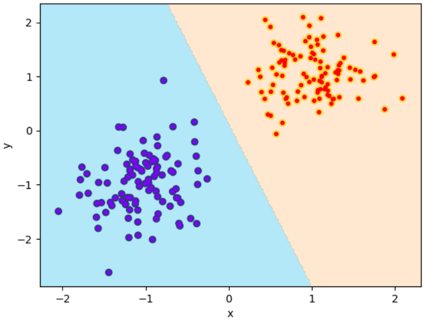
Hands-On Machine Learning with Scikit-Learn and TensorFlow

Aurélien Géron

https://github.com/ageron/handson-ml/

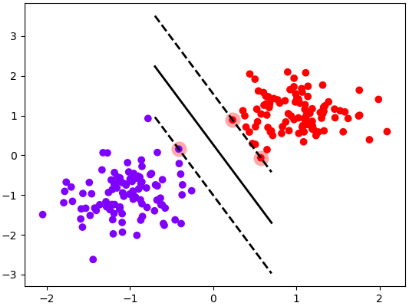
Outline

Definition



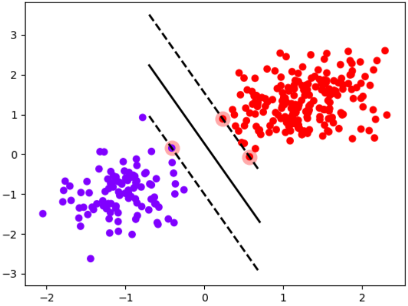
- SVMs are binary classifiers, i.e. each sample can be member of either two categories
- In its simplest form, looking for linear separation

Principle idea



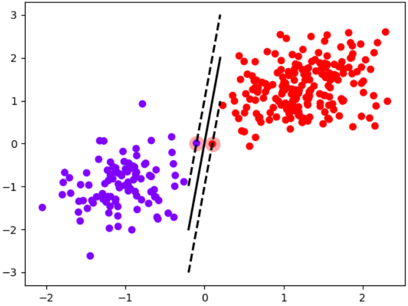
- Find line that separates and keeps largest possible distance
- Data points closest to line determine parameters and are called Support Vectors

Robustness – Part I



- Additional points away from decision boundary have no influence on result
- Same Support Vectors as before

Robustness – Part II



- Small changes near the decision boundary can have drastic impact on the result

Setup

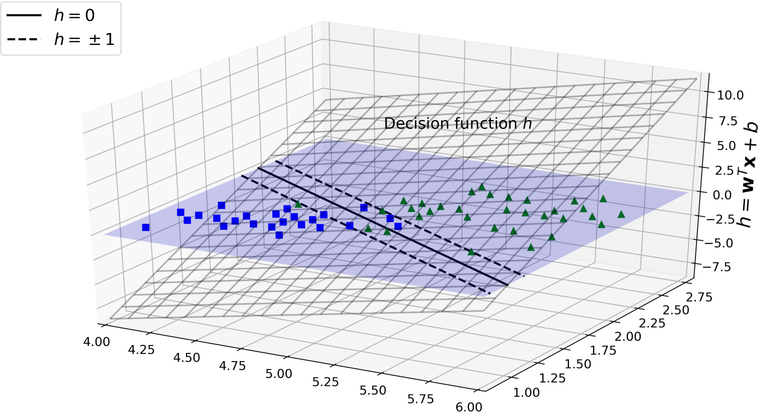
We have a set of N training data points

$$(\mathbf{x}_i, y_i), \quad i = 1 \dots N, \quad y_i \in \{-1, 1\},$$

want to find hyper-plane with parameters $\mathbf{w}$, b so that

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x} + b &\geq 1 \quad \text{for } y_i = 1 && (1) \\ \mathbf{w} \cdot \mathbf{x} + b &\leq -1 \quad \text{for } y_i = -1 && (2) \end{aligned}$$

Graphical illustration



We also want to have the maximum margin width (distance between dashed lines), where $h = \pm 1$.

Question 1

Question

How can we combine

$$\mathbf{w} \cdot \mathbf{x} + b \geq 1 \quad \text{for } y_i = 1 \tag{3}$$

$$\mathbf{w} \cdot \mathbf{x} + b \leq -1 \quad \text{for } y_i = -1 \tag{4}$$

into a single equation?

Question 1

Question

How can we combine

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x} + b &\geq 1 \quad \text{for } y_i = 1 && (3) \\ \mathbf{w} \cdot \mathbf{x} + b &\leq -1 \quad \text{for } y_i = -1 && (4) \end{aligned}$$

into a single equation?

Answer

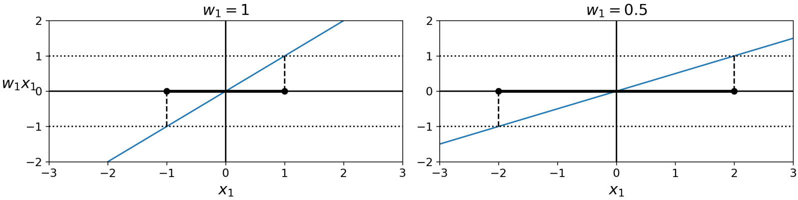
$$y_i\,(\mathbf{w} \cdot \mathbf{x} + b) \geq 1$$

Question 2

Question

How can we make the margin as wide as possible?

Question 2



Question

How can we make the margin as wide as possible?

Answer

Minimize $\|\mathbf{w}\|$.

Problem formulation

We need to minimize

$$\|\mathbf{w}\| \text{ subject to } y_i\,(\mathbf{w}\cdot\mathbf{x}+b) \geq 1$$

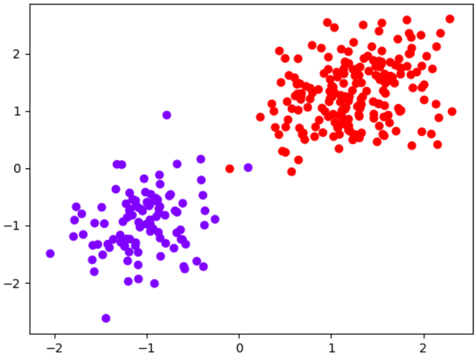
This is a non-linear optimization problem. Solved by quadratic programming.

$$\mathbf{p} = (\mathbf{w}, b) \tag{5}$$

$$\frac{1}{2}\mathbf{pHp} + \mathbf{fp}^T \to \min \tag{6}$$

$$\mathbf{Ap} \leq \mathbf{b} \tag{7}$$

Soft margin SVM



So far we have assumed that our classes are linearly seperable. Cannot solve above situation.

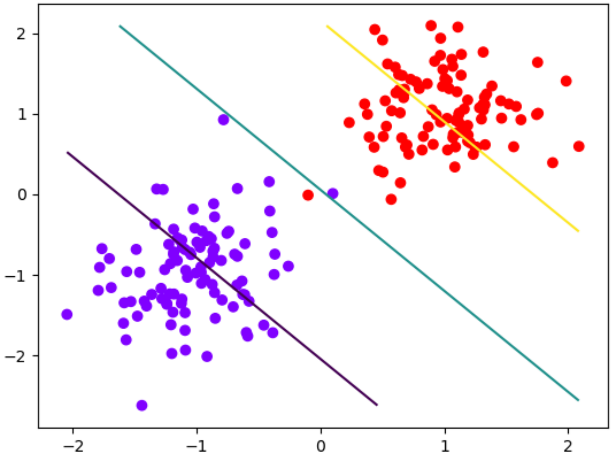
Slack variables

Introduce slack variables ζ_i and modify minimization problem

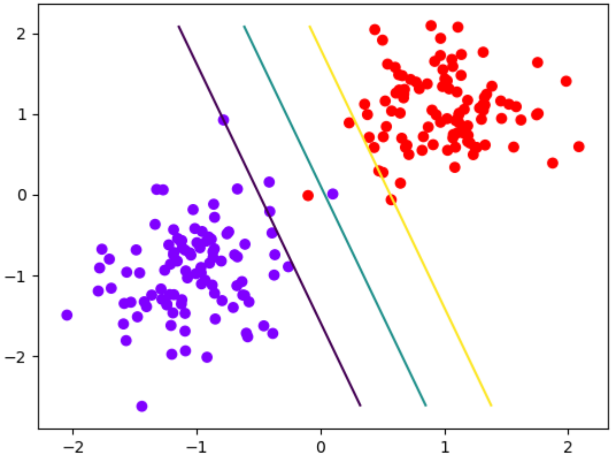
$$\|\mathbf{w}\| + C\sum_{i=1}^{m}\zeta_i \text{ subject to } y_i\left(\mathbf{w}\cdot\mathbf{x}+b\right) \geq 1-\zeta_i$$

- Also a quadratic programming problem, extra term compared to hard-marging
- C balances between margin width (large for small C) and boundary violations (small for large C)

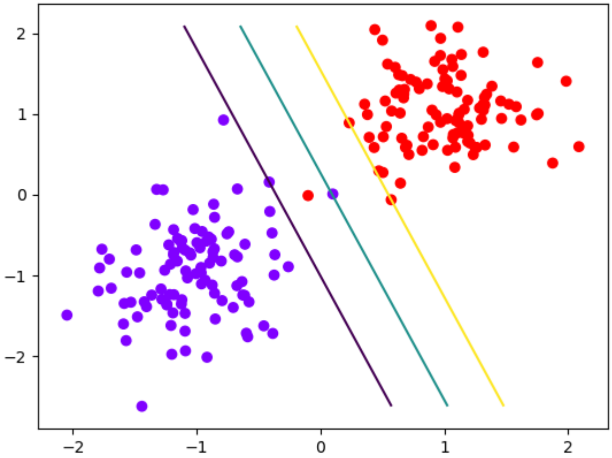
C =0.01



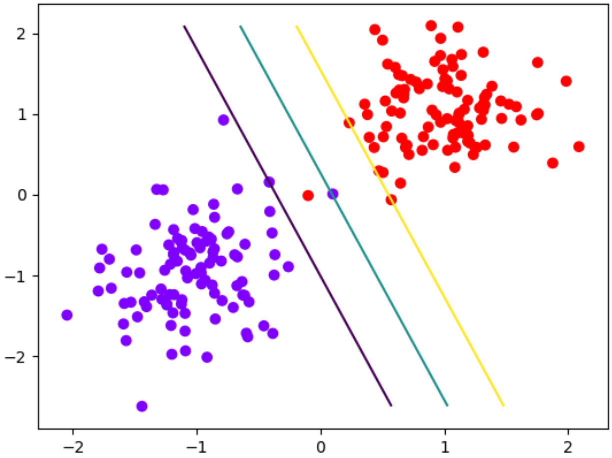
C=1



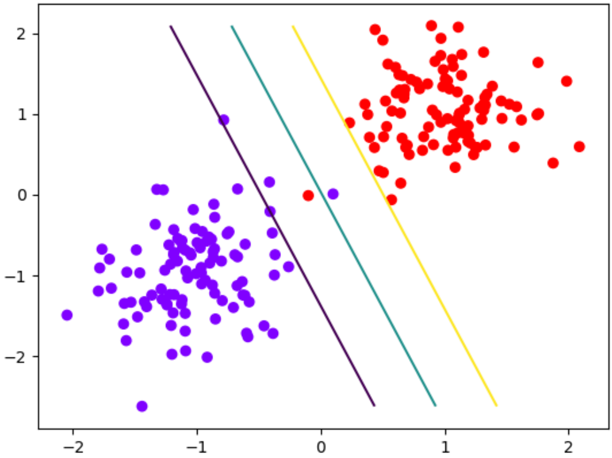
C=10



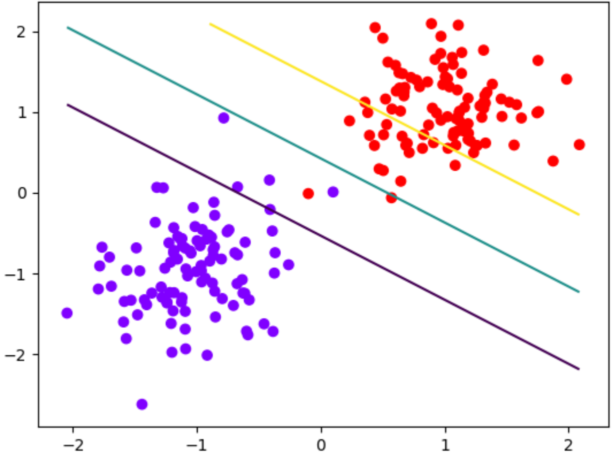
C=100



C=1,000



C=10,000



Support vector regression

Now have two slack variables ζ_i, $\zeta_i^\star$ and user-defined parameters ϵ, C

$$\|\mathbf{w}\| + C\sum_{i=1}^{m}(\zeta_i + \zeta_i^\star) \text{ subject to} \tag{8}$$

$$y_i - (\mathbf{w}\cdot\mathbf{x} + b) \leq \epsilon + \zeta_i \tag{9}$$

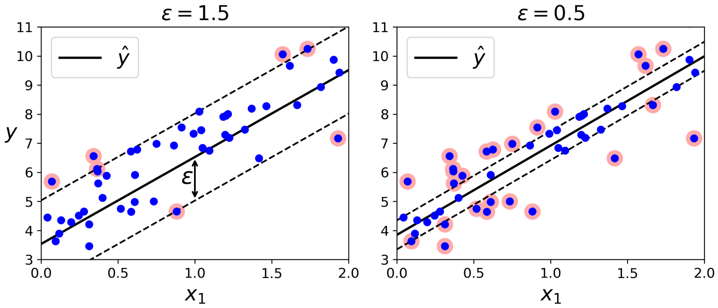
$$(\mathbf{w}\cdot\mathbf{x} + b) - y_i \leq \epsilon + \zeta_i^\star \tag{10}$$

$$\zeta_i \geq 0 \tag{11}$$

$$\zeta_i^\star \geq 0 \tag{12}$$

ϵ defines the width of the margin.

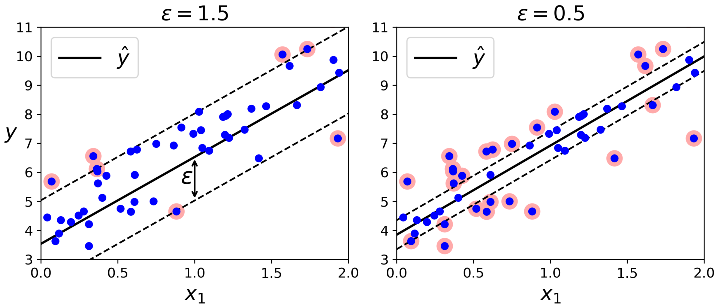
Graphical representation



Question

What is the role of $||\mathbf{w}||$ in the minimization?

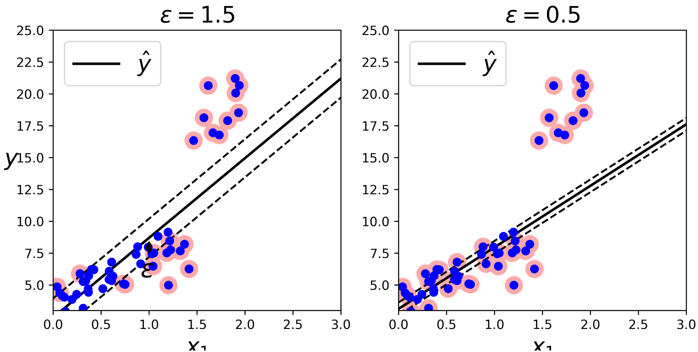
Graphical representation



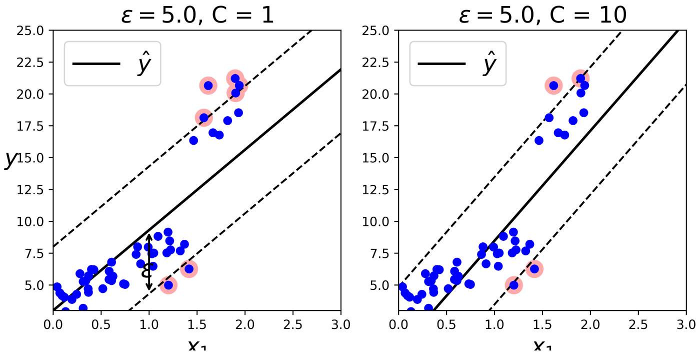
Answer

Minimize the slope of the line/plane.

Impact of ϵ



Impact of C



$\varepsilon = 5.0$, C = 1
$\varepsilon = 5.0$, C = 10
$\hat{y}$
ε
y
x_1
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Summary

- Support vector machines can be used for classification and regression
- Solution obtained by solving a constrained optimization problem
- Hard-margin SVM requires clearly separated clusters for classification
- Soft-margin allows for violations
- SVMs have user tunable regularization parameters (C,ϵ) with potentially strong impact on results.