Contents

Contents

Contents

Contents

Contents

Contents

Contents

Outline

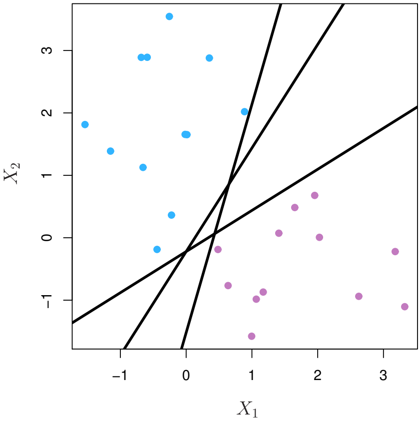
1. Support Vector Machines
   - Motivations
   - Hyperplanes
   - Maximal Marging classifier
   - Support Vector classifier
   - Kernel trick
   - Other topics
   - SVM or Logistic Regression

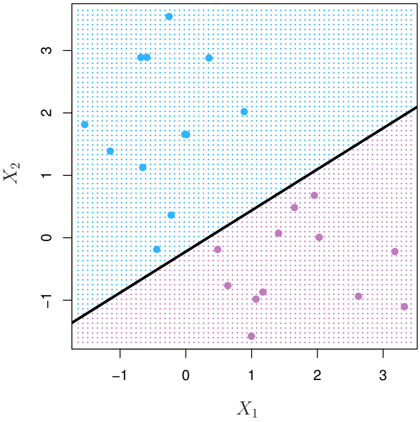
In the classification context, try to separate a two-class population by a plane or an hyperplane ($p > 2$)

- Maximal Margin classifier, which is translated in a convex optimization problem ;
- When the frontier is clearly non linear, we can use the kernel trick : convert the problem in another (bigger) space where the frontier will be more regular.
- History, one of most successful approach at the end of 90's (V.Vapnik)

Figures Hyperplanes

Outline

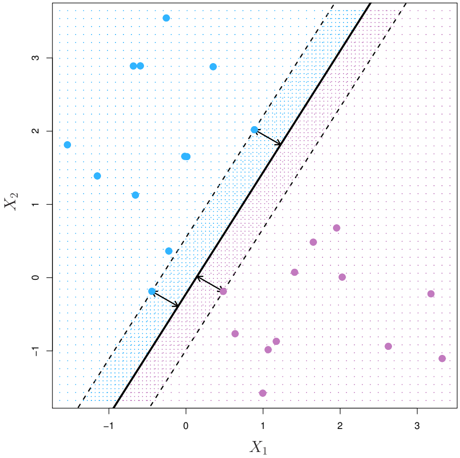




- On the left, three separating hyperplanes
- On the right, a separating hyperplane with the associated regions

Outline

1. Support Vector Machines
   - Motivations
   - Hyperplanes
   - Maximal Marging classifier
   - Support Vector classifier
   - Kernel trick
   - Other topics
   - SVM or Logistic Regression

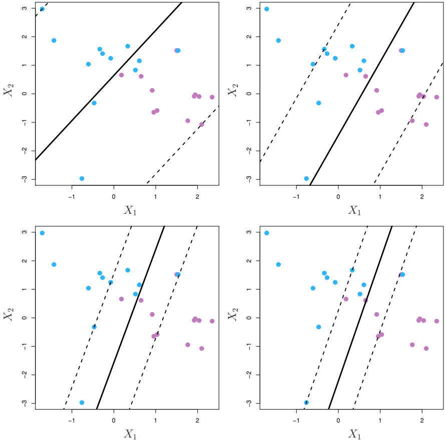


- Among all separating hyperplanes, find the one that makes the biggest gap or margin between the two classes.

Constrained optimization problem

- The preceding problem could be translated in :
  $maximize_{(\beta_1,\ldots,\beta_p)} M$
  subject to $\sum_{i=1}^{p} \beta_i^2 = 1$ and
  $y_i(\beta_0 + \beta_1 x_{i1} + \ldots \beta_p x_{ip}) \geq M$ for all $i = 1 \ldots p$
- This can be rephrased as a convex quadratic program, and solved efficiently by the function svm() in R-package e1071.

Outline



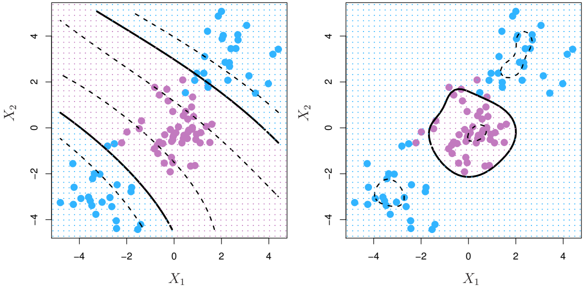
- Taking in account the case when there is no separating hyperplane, we relax the constraint

Constrained optimization problem with relaxation

- The optimization problem becomes :
  $maximize_{(\beta_1,\ldots,\beta_p)} M$ subject to $\sum_{i=1}^{p} \beta_i^2 = 1$ and $y_i(\beta_0 + \beta_1 x_{i1} + \ldots\beta_p x_{ip}) \geq M(1 - \epsilon_i)$ for all $i = 1\ldots p$ and $\epsilon_i > 0$ and $\sum_{i=1}^{p} \epsilon_i \leq C$
- C is a regularization parameter

Outline

1. Support Vector Machines
   - Motivations
   - Hyperplanes
   - Maximal Marging classifier
   - Support Vector classifier
   - Kernel trick
   - Other topics
   - SVM or Logistic Regression



- This is a basis transformation which take in account non-linearities :
    1. Polynomial expansion (left)
    2. New space, new variables (right)

Kernel trick is a change of base

Approaches based on Support Vector Machine to Classification of Remote Sensing Data 337

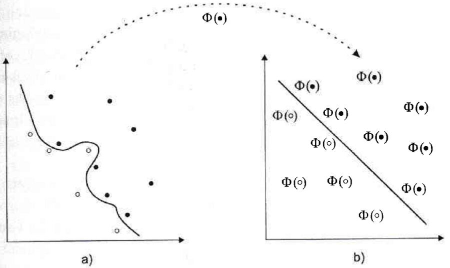


Fig. 2. Transformation of the input data by means of a kernel function into a high dimension feature space. a) Input feature space; b) kernel induced high dimensional feature space.

- This is a basis transformation which take in account non-linearities

Kernel trick is a change of base

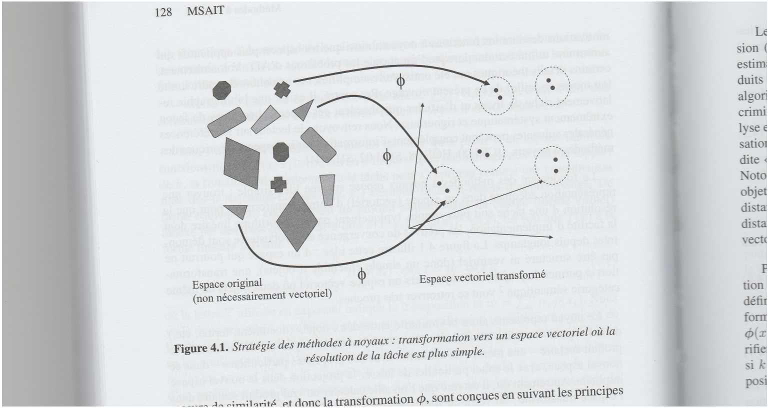


Figure 4.1. *Stratégie des méthodes à noyaux : transformation vers un espace vectoriel où la résolution de la tâche est plus simple.*

et donc la transformation ϕ, sont conçues en suivant les principes

- This is a basis transformation which take in account non-linearities

Outline

1. Support Vector Machines
   - Motivations
   - Hyperplanes
   - Maximal Marging classifier
   - Support Vector classifier
   - Kernel trick
   - Other topics
   - SVM or Logistic Regression

- Compare SVM and Neural Network (two layers)
- More than two classes classification : OVA (One versus All), OVO (One versus One)

Outline

Which to use : SVM or Logistic Regression ?

- When classes are (nearly) separable, SVM does better than LR. So does LDA.
- When not, LR (with ridge penalty) and SVM very similar.
- If you wish to estimate probabilities, LR is the choice. Computation of a confident interval (CI).
- For nonlinear boundaries, kernel SVMs are popular. Can use kernels with LR and LDA as well, but computations are more expensive.