04_SVM

1. Complexity
Adaptation of the Complexity
VC-Dimension
2. SVM: Supported Vector Machine
Optimization Model: A Constrained Optimization Model
Problems and Solution: Class Overlap and Nonlinearity
Class Overlap Solution: SVM with Slacks
Nonlinearity Solution: Kernel Trick
3. One-Class Classification
Background
From Classifiers to Outlier Detection
Summary

1. Complexity

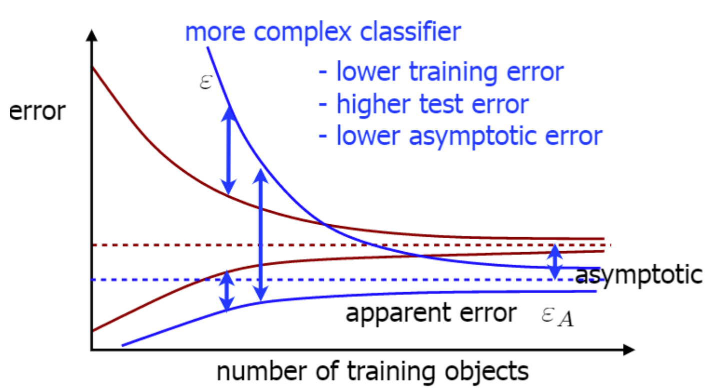
Definition:

The complexity of a classifier indicates **the ability to fit to any data distribution.**

- A simple classifier fits to only a few specific data distributions
- A complex classifier fits to almost all data distributions

Adaptation of the Complexity

A learning curve indicate the complexity of the model and the number of training objects is shown in Figure.



So based on that, we should understand that we should **Choose the complexity according to the available training set size**.

VC-Dimension

Definition:

The largest number h that a given test can be separated in 2^h possible ways.

Property:

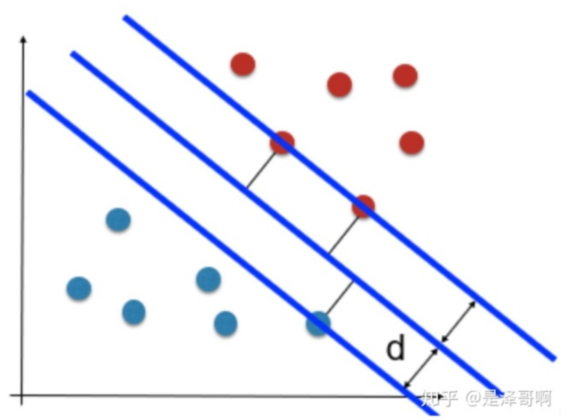
The detailed of VC-dimension will not be introduced, however we should know:

- Unfortunately, only for a very few classifiers the VC-dimension is known
- Fortunately, when you know h of a classifier, you can bound the true error of the classifier
- When h is small, the true error is close to the apparent error

2. SVM: Supported Vector Machine

Idea

SVM try to find an **optimal decision boundary**. The boundary should be the **furthest boundary to the nearest samples** in two classes. I.e., it tries to maximize the margin. A example figure is shown in the Figure.



Optimization Model: A Constrained Optimization Model

Based on algebra, the distance of the **supported vector** to the decision boundary can be represented by $d = \frac{|w^T x + b|}{\|w\|}$. Then we will have $||\mathbf{w}||^2 = \frac{2}{d^2}$. Then the problem can be represented by:

$$\min \frac{1}{2}||w||^2$$
$$y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq +1 \quad \text{for all } i$$

We can use **Lagrange Multiplier** Method to solve it.

$$\mathcal{L}(\mathbf{w}, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i \alpha_i \left(y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) - 1\right)$$

By using derivation, we can then come to the result:

$$\begin{aligned}\mathcal{L}(\mathbf{w}, \alpha) &= \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i \alpha_i \left(y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) - 1\right) \\ \frac{\partial \mathcal{L}(\mathbf{w}, \alpha)}{\partial \mathbf{w}} &= \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \\ \frac{\partial L}{\partial b} &= \sum_{i=1}^{n} \alpha_i y_i = 0\end{aligned}$$

Then the next problem is to determine α, take the representation of w into the problem and based on our target to maximum the distance to boundary, we will generate a new problem:

$$\begin{gathered}\max_\alpha \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j}^{N} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \\ \alpha_i \geq 0 \quad \forall i \\ \sum_{i=1}^{N} \alpha_i y_i = 0 \\ \mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i\end{gathered}$$

This is a standard **quadratic optimization problem**.

Notes: The SVM is determined by objects, not features. We do not "learn" the distribution of a class on a given features. When classifying, we **only use the value of feature of given project**.

Problems and Solution: Class Overlap and Nonlinearity

The **limitation** of standard SVM is:

- The data should be **separable**
- The decision boundary is **linear**

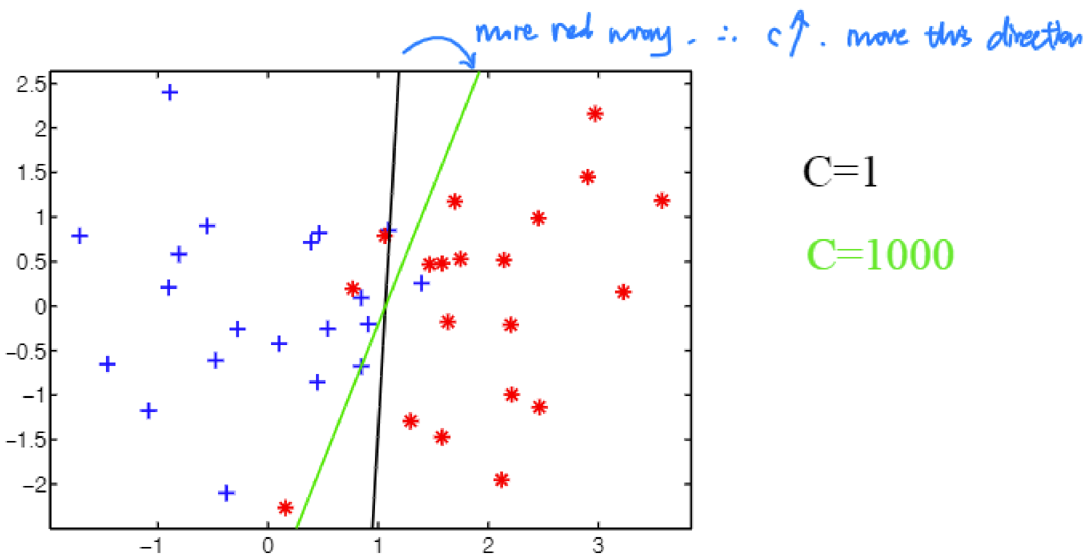
Class Overlap Solution: SVM with Slacks

Class Overlap Solution: SVM with Slacks

To solve the class overlap problem, we can introduce the **slacks Variables**

$$
\begin{aligned}
&\min \|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\
&\mathbf{w}^T\mathbf{x}_i + b \geq +1 - \xi_i, \qquad \text{for } y_i = +1 \\
&\mathbf{w}^T\mathbf{x}_i + b \leq -1 + \xi_i, \qquad \text{for } y_i = -1 \\
&\quad \xi_i \geq 0 \quad \forall i
\end{aligned}
$$

Then there will be a **tradeoff** of parameter C, parameter C weighs the contributions between the training error and the structural error.



Nonlinearity Solution: Kernel Trick

For solving the problem of nonlinear decision boundary, we can use **kernel method**. Currently in standard SVM, we can observe that **all operations are on inner products between objects**.

Introduction: Kernel Trick

The idea to replace all inner products by a single function, the kernel function K.

Kernel Trick Optimization Model

The kernel trick implicitly **maps the data to a high dimensional feature space**. Then the problem will become:

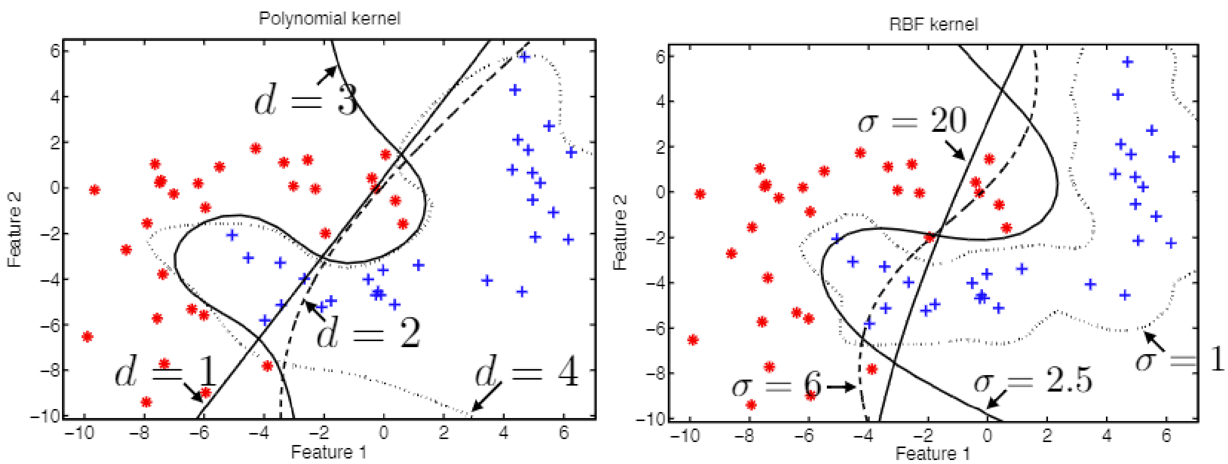
$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j}^{N} y_i y_j \alpha_i \alpha_j K\left(\mathbf{x}_i, \mathbf{x}_j\right)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad \forall i$$

$$f(\mathbf{z}) = \sum_{i=1}^{N} \alpha_i y_i K\left(\mathbf{x}_i, \mathbf{z}\right) + b$$

<u>Kernel Types</u>

There are several examples of kernel function in the Figure



$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^d \qquad K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right)$$

There are two points in these kernel functions to be noticed:

- In the Polynomial Kernel, the $+1$ part is in order to add polynomials with lower order element
- In RBF kernel, if we has larger σ, which will make the boundary more smoother.

3. One-Class Classification

Background

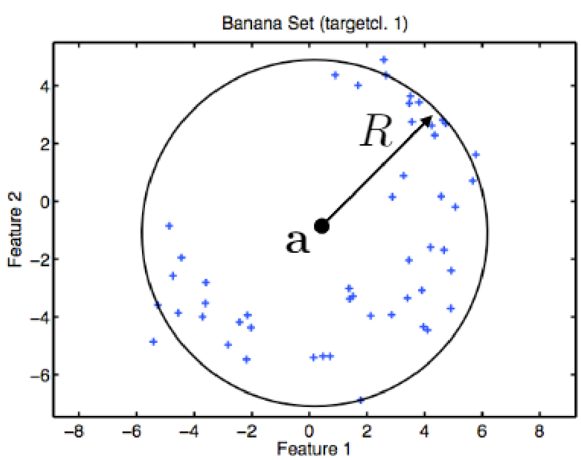
<u>One-class Classification problem</u> is classical in **abnormal detection**. Because abnormal conditions (outlier class) are rare and hard to sample reliably.

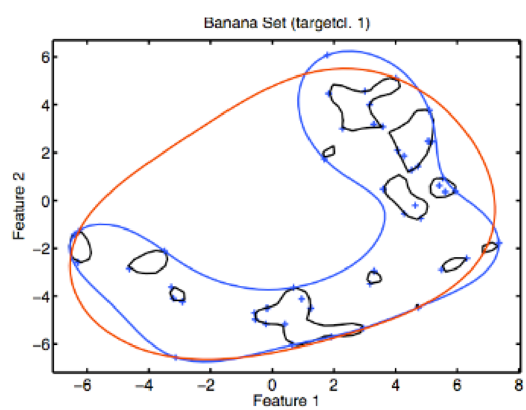
From Classifiers to Outlier Detection

One classical way to deal with that is to build a hypersphere around the target class. We can call it **Support Vector Data Description**.

Besides we can also use a kernel functions. A example in Figure the figure

	Support Vector cl.	Support vector DD
model	hyperplane $\mathbf{w}, b$	hypersphere $\mathbf{a}, R$
complexity	$\Vert\mathbf{w}\Vert^2$	R^2
error	$\Vert\mathbf{w}\Vert^2 + C\sum_i \xi_i$	$R^2 + C\sum_i \xi_i$
SVs	objects on the plane	objects on the sphere
slacks	objects on the wrong side of the plane	objects outside the sphere





Radial basis (Gaussian)

$$K(\mathbf{z}, \mathbf{x}) = \exp\left(-\frac{\Vert\mathbf{x} - \mathbf{z}\Vert^2}{\sigma^2}\right)$$

Summary