

Kernel PCA for Novelty Detection

Heiko Hoffman (2005)

Rodríguez Jeremías

November 16, 2017

Table of Contents

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades
- 5 Experimentos
- 6 Conclusión

Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades
- 5 Experimentos
- 6 Conclusión

Problema de clasificación

Tenemos un conjunto de puntos, etiquetados con **dos o más clases**:

$$X = \{(x_i, y_i) \mid x_i \in R^D, y_i \in \{-1, 1\}, i = 1 \dots N\}$$

Con ellos, entrenamos una máquina que clasifica nuevas instancias:

$$h : R^D \mapsto \{-1, 1\}$$

Estudiamos varios métodos:

- Support vector machines
- Árboles de decisión
- Redes neuronales

Novelty Detection: Motivación

En muchos problemas, tenemos una gran cantidad de datos de **una sola clase**. Conseguir datos de la otra puede ser muy complejo, costoso o incluso imposible.

- Detección de transacciones fraudulentas con tarjeta de crédito.
- Diferenciar tejidos sanos de aquellos que presentan cancer maligno.
- Catálogos astronómicos.
- Detección de fallas en sistemas industriales.

En estos problemas, es relativamente sencillo reunir datos en donde las situaciones están OK, pero recolectar datos de situaciones anómalas puede ser muy complejo.

Novelty Detection: Definición

Deteccion de novedades

Tenemos datos correspondientes a **una sola clase**:

$$X = \{x_i \mid x_i \in R^D, i = 1 \dots N\}$$

Y entrenamos una máquina que reconoce si una nueva instancia pertenece a la clase conocida, o es una novedad:

$$z : R^D \mapsto R$$

Dada un nuevo punto p , se le asigna un **novelty score** $z(p)$.

Se define un umbral k , tal que si $z(p) > k$, p es considerado una novedad.

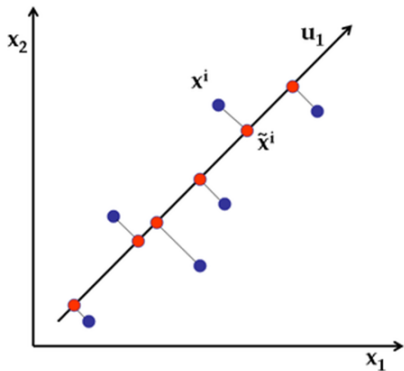
Este paper presenta un nuevo método, kernel PCA, que mejora los existentes para abordar este tipo de problemas.

Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades
- 5 Experimentos
- 6 Conclusión

Principal Components Analysis (PCA)

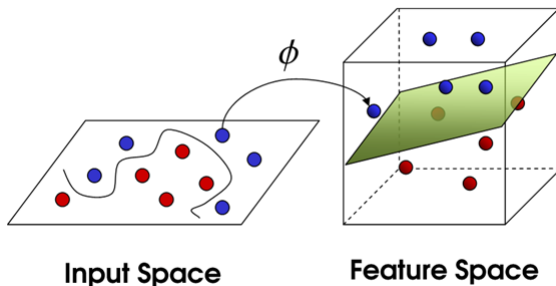
Dado un dataset X con n puntos y p dimensiones, queremos encontrar una representación de estos mismos datos **en un espacio de menor dimensión** de tal forma que conserven su estructura.



- Buscamos la dirección que minimiza la varianza o que maximiza los **errores de reconstrucción**
- La segunda dirección principal es aquella ortogonal a la primera que maximiza la varianza, y así sucesivamente.
- Proyectamos los puntos originales sobre ese subespacio.
- PCA hace un giro de los ejes

SVM y el truco del kernel

- Para resolver problemas de clasificación estudiamos Support Vector Machines, que buscan separar las clases mediante un hiperplano que maximice el margen.
- Si los datos no son linealmente separables, se utilizan distintos kernels (polinomial, RBF)



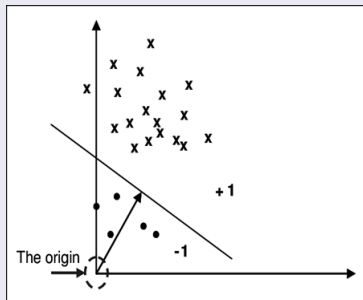
- No necesitamos conocer ϕ

Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection**
- 4 Kernel PCA para detección de novedades
- 5 Experimentos
- 6 Conclusión

Métodos previos: Usando SVM

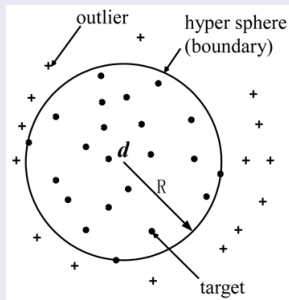
One-Class SVM



Separa todos los datos de training del origen en feature space, con un hiperplano lo mas alejado posible del origen.

Problema: Encierran demasiado espacio.

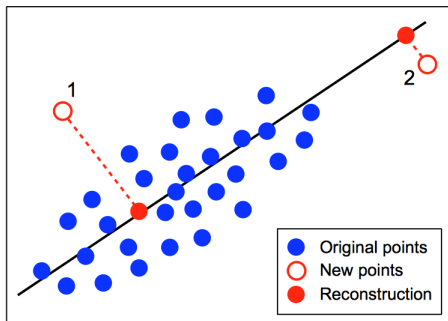
SVDD



Circunscribe los puntos de training en una hiperesfera tan ajustada como sea posible, en feature space.

Métodos previos: Usando modelado de datos

En datasets linealmente separables, se puede aplicar PCA y utilizar el error de reconstrucción como medida de novedad.



En datasets no lineales, los métodos hasta el momento desarrollados son sensibles a inicialización y a mínimos locales.

Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades**
- 5 Experimentos
- 6 Conclusión

Método Kernel PCA: Idea general

Hoffman propone combinar PCA y kernels para obtener un método de detección de novedades sin los problemas mencionados. Partimos de un conjunto de training con datos de una sola clase:

$$X = \{x_i \mid x_i \in R^D, i = 1 \dots N\}$$

Mapeamos los puntos a un espacio de mayor dimensionalidad:

$$x_i \mapsto \phi(x_i)$$

En este espacio realizamos PCA standard. Luego, utilizamos como scores los errores de reconstrucción.

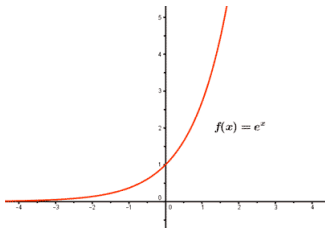
- No necesitamos la transformación completa. Hoffman propone una forma de realizar el cálculo de la PCA y de los scores usando solo un kernel.
- Un parámetro libre sera el número de componentes principales q a conservar luego de aplicar PCA.

Método Kernel PCA: El kernel RBF

En este tipo de problemas, el kernel que se utiliza por tener mayor eficacia es Radial Basis Function (RBF):

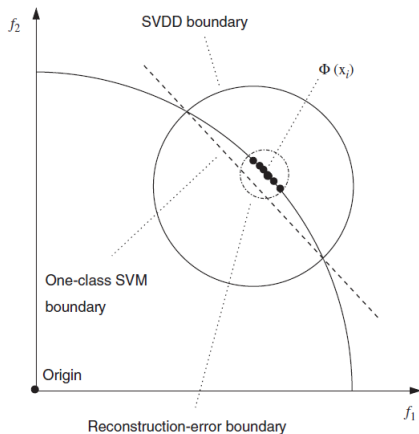
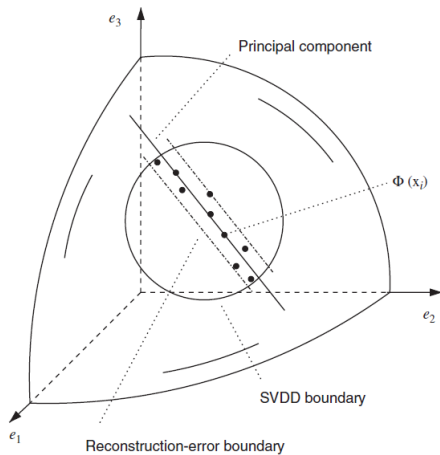
$$K(x, y) = e^{-\frac{\|x-x'\|^2}{2\gamma^2}} \approx \sum_{n=0}^{+\infty} \frac{(x \cdot y)^n}{\gamma^n n!}$$

Donde γ es un parámetro libre y $\|\cdot\|$ es la norma euclídea.



- Siempre retorna valores en $(0,1]$
- Si $x=y$, $K(x,y)=1$. Cuanto más lejos está x de y , $K(x,y)$ asume valores más cercanos a 0. Por lo tanto, $K(x,y)$ es una medida de similitud.
- El feature space correspondiente tiene un **número infinito de dimensiones!**
- Todos los puntos están ubicados sobre una hipersfera de radio 1.

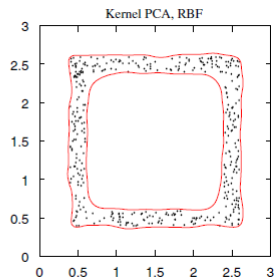
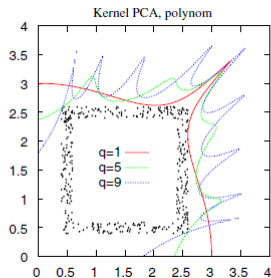
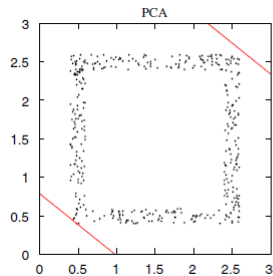
Método Kernel PCA usando RBF



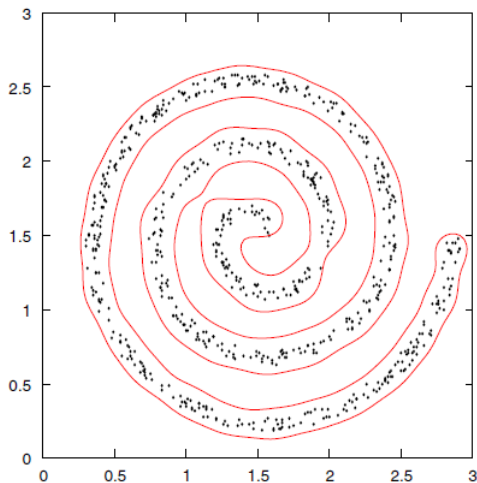
Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades
- 5 Experimentos**
- 6 Conclusión

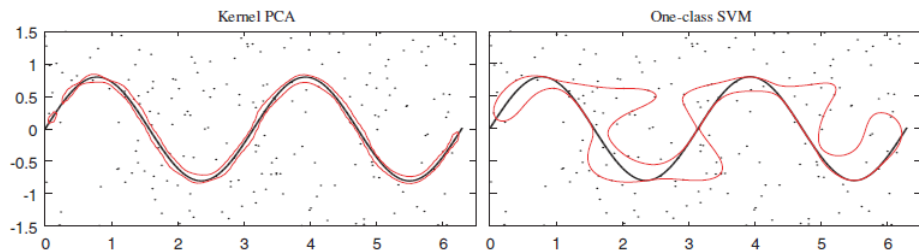
Comparación de kernels



Kernel PCA con RBF



Kernel PCA vs One class SVM: Ruido



- Los métodos tienen sus parámetros optimizados
- Los parámetros de Kernel PCA (con RBF) son q y γ

Contenido

- 1 Novelty Detection
- 2 Repaso de contenidos previos
- 3 Métodos previos para Novelty Detection
- 4 Kernel PCA para detección de novedades
- 5 Experimentos
- 6 Conclusión**

- Se presentó el problema de detección de novedades
- Se presentó el método kernel PCA para detección de novedades.
- Los datos de training son mapeados a un espacio de infinitas dimensiones, y luego proyectados a un subespacio utilizando PCA.
- El error de reconstrucción de una nueva instancia con respecto a este subespacio fue utilizado como medida de novedad.
- El método demostró alta performance con respecto a los métodos previos, tanto en datasets sintéticos como reales.