

Introducción a Machine Learning

Otros modelos de clasificación

Ronald Cárdenas Acosta

Agosto, 2016

Outline

- 1 Support Vector Machines
 - Función Objetivo
 - Kernel Trick

- 2 Árboles de Decisión (Decision Trees)
 - Definición
 - Random Forest

Outline

- 1 Support Vector Machines
 - Función Objetivo
 - Kernel Trick

- 2 Árboles de Decisión (Decision Trees)
 - Definición
 - Random Forest

Support Vector Machines

- Usadas para clasificación, regresión y detección de anomalías.
- Capaz de modelar dependencias no lineales mediante el uso de Kernels
- Eficiente en memoria: usa las muestras más cercanas a la línea de decisión para la predicción
- Desventaja: si el número de características es mayor al número de muestras, se tendrá bajo rendimiento

SVM: Función Objetivo

$$L(w) = C * \sum_{i=1}^N [y^i \cdot \text{Cost}_1(w * x^i) + (1 - y^i) \cdot \text{Cost}_0(w * x^i)] + \frac{1}{2} \sum_{j=1}^M w_j^2$$

- Para SVM Lineal, se plantea una minimización con restricciones

$$\begin{aligned} & \min \|w\| \\ & \text{restringido a} \\ & y^i(x^i * w + b) \geq 1 \end{aligned}$$

Outline

1 Support Vector Machines

- Función Objetivo
- Kernel Trick

2 Árboles de Decisión (Decision Trees)

- Definición
- Random Forest

Kernels

- Mapea la entrada a un espacio de alta dimensión en el que la data sea linealmente separable
- Para el caso lineal se reemplaza $x * w$ por $K(x)$
- Kernels más usados:
 - Lineal
 - Polinomial
 - Gaussiano (RBF)
 - Tangente hiperbólica

Función Kernel

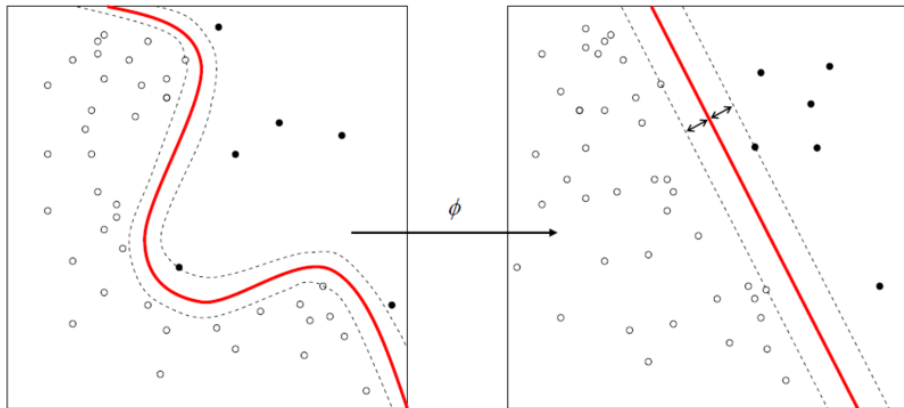


Figure: Mapeo de función Kernel

Outline

- 1 Support Vector Machines
 - Función Objetivo
 - Kernel Trick
- 2 Árboles de Decisión (Decision Trees)
 - Definición
 - Random Forest

Árboles de Decisión

- Considerados los modelos más humanamente interpretables
- Usados para clasificación y regresión
- Deducen reglas de discriminación jerárquicas entre las características, formando un árbol
- La raíz representa la decisión inicial y las hojas las clases predecidas
- Tienen a sobre-ajustar la data

Árboles de Decisión

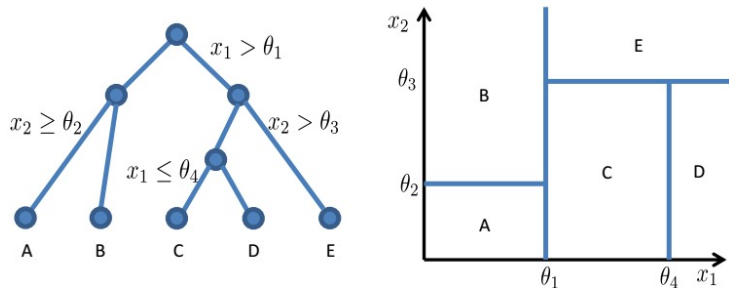


Figure: Árbol de decisión

Outline

- 1 Support Vector Machines
 - Función Objetivo
 - Kernel Trick
- 2 Árboles de Decisión (Decision Trees)
 - Definición
 - Random Forest

Random Forest

- Generalización de árboles de decisión
- Se dividen las características en grupos disjuntos y se entrena un árbol en cada grupo
- La decisión final se hace de acuerdo a un promedio del resultado de cada árbol
- Útil cuando se tienen características dependientes
- Ejemplo de *Ensemble model*

Random Forest

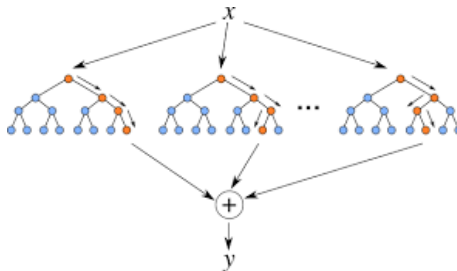


Figure: Random Forest: un conjunto de árboles