

# Support Vector Machine

MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA

---

Pablo Alfaro Goicoechea

Carlos Morales Aguilera

Carlos Santiago Sánchez Muñoz

26 de Diciembre de 2020

Tratamiento Inteligente de Datos

*E.T.S. de Ingenierías Informática y de Telecomunicación*



**UNIVERSIDAD  
DE GRANADA**



Presentación

¿Qué son las SVM?

Fundamentos

- Principios

- Fronteras de decisión

- Truco del kernel

Aplicaciones

- Problemas tipo

- Aplicaciones actuales

Ejemplos de implementación

- Dataset Iris

- Función lineal de 2  
dimensiones

Comparativa entre algoritmos

- Comparativa entre algoritmos

- Elección de algoritmos

Conclusiones

# Presentación

---

## Objetivos:

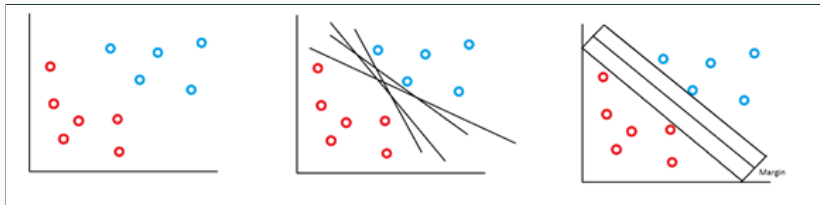
- Conocer los principios matemáticos de SVM.
- Comprender los casos relevantes para el uso de SVM.
- Entender las diferencias respecto a otros algoritmos de clasificación.
- Realizar algún ejemplo de implementación.
- Obtener unas conclusiones sobre sus ventajas y utilidad.

# ¿Qué son las SVM?

---

# ¿Qué son las SVM?

- Son un tipo de modelo de clasificación lineal supervisada
- Se centran en mejorar la generalización. Esto quiere decir que será más robusto a la hora de clasificar nuevos ejemplos que no se parezcan tanto a los de entrenamiento.



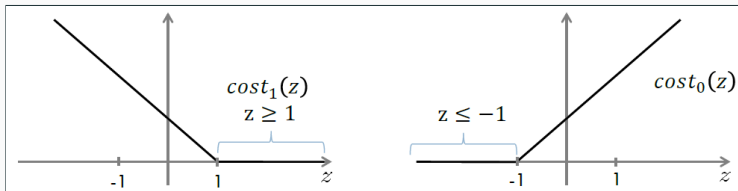
Ejemplo de SVM.

# Fundamentos

---

# Principios

- Para encontrar la mejor frontera de decisión hay que clasificar las posibles de alguna manera.
- Se utiliza una función de coste que evalúa la certeza que tiene el modelo al clasificar cada ejemplo.
- Una función que se suele utilizar es la siguiente:

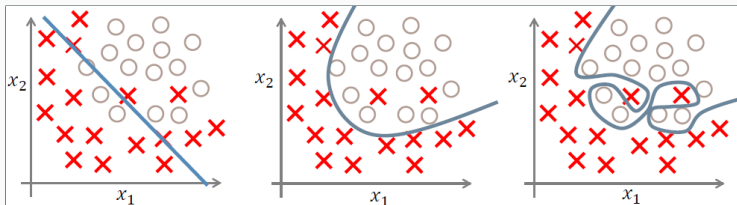


Función de coste.



# Regularización

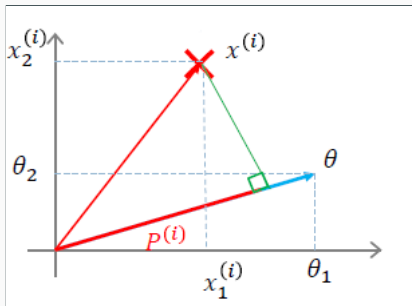
- El parámetro de regularización indica la penalización de clasificar mal un ejemplo.
- Elegir el bien valor para este parámetro es fundamental.
- Es la diferencia entre tener una varianza alta, sobreaprender o tener una buena frontera de decisión.



Posibles fronteras de decisión.

# Fronteras de decisión

- Asumiendo una simplificación para un conjunto bidimensional de  $\theta_0 = 0$  y  $n = 2$ .
- Utilizando la teoría de los productos escalares,  $\theta^T x^{(i)}$  se podría ver de esta manera:



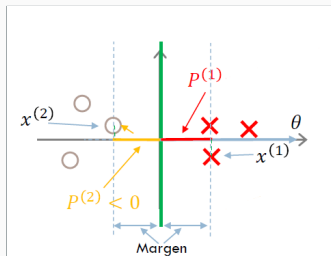
$$\theta^T x^{(i)}.$$

# Fronteras de decisión

- Viendo lo anterior, se puede concluir lo siguiente:

$$\theta^T x^{(i)} = P^{(i)} \|\theta\|.$$

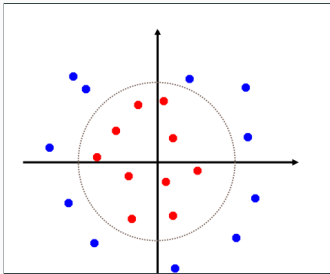
- Por tanto se busca que las proyecciones de los ejemplos sobre  $\theta$  sean lo más grandes posibles.
- Esto es lo que provoca que la SVM busque márgenes grandes. Haciendo esto, la SVM puede obtener una norma de  $\theta$  menor.



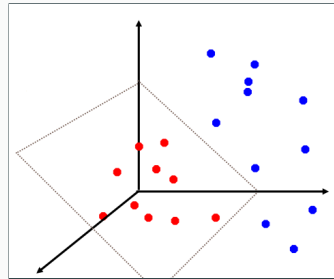
Cálculo del margen.

# Truco del kernel

- Cuando la frontera que hay que establecer no es lineal se recurre al Kernel.
- De esta manera se mapea el espacio de características a uno mucho mayor.
- El hiperplano de margen máximo se busca en este nuevo espacio.



Espacio original.



Nuevo espacio.

- Gracias al uso de las funciones kernel, se evita que el coste computacional aumente.
- El resultado de la función kernel es el mismo que el de trasladar cada ejemplo por separado a un espacio de características mucho mayor y luego calcular su producto escalar.
- En realidad no necesitamos calcular el nuevo conjunto de características para cada ejemplo.

# Aplicaciones

---

SVM funciona bien para problemas de clasificación y de regresión.

## **Ventajas de SVM**

- Adecuado cuando existe margen de separación de las clases.
- Efectivo en espacios de dimensión alta.
- Número de dimensiones es mayor al número de ejemplos.
- Eficiente en memoria.

## **Desventajas de SVM**

- No es adecuado para grandes conjuntos de datos.
- No da buenos resultados cuando hay ruido o solapamiento.
- No hay una explicación probabilística para la clasificación.

- El incremento de la esperanza de vida hace más presentes enfermedades de la tercera edad como la Enfermedad del Parkinson(EP) y el Temblor Esencial (TE).
- Potencial relación hace difícil el diagnóstico (error del 40 %).
- Base de datos extraída de un dispositivo háptico consensuado con especialistas y expertos.
- Utilizando SVM se han conseguido ratios de acierto cercanos al 100 %.



- Reconocimiento de Locutor. Inconveniente: variabilidad del canal y ruido.
- Interés: herramienta útil al pinchar teléfonos.
- Paradigma SVM ha demostrado ser superior a otras estrategias de fusión basada en aprendizaje (redes neuronales).
- Base de datos con más de 9h de habla durante 2 meses.
- SVM mejora hasta en un 40 % las técnicas existentes.

- Los rasgos faciales son una característica biométrica única.
- Se ha utilizado un sistema de decisión SVM multiclase, tanto lineal como no lineal.
- SVM uno vs resto (1 vs R).
- SVM-DAG (Directed Acyclic Graph).
- El acierto de estos modelos suele estar sobre el 90 %.

- Jerarquización de documentos basada en la estadística de las palabras presentes en dicho documento y la consulta.
- Colección MEDLINE contiene 1033 documentos y se hacen 30 consultas.
- Mediante SVM se ha construido un clasificador.
- ¿Cómo? Construcción de vectores usando el documento, las palabras ordenadas de manera específica dando un valor extra a los valores de la consulta. Los valores que sobran del reordenamiento se eliminan.

## **Ejemplos de implementación**

---

## Implementación

- Se comprueban diferentes algoritmos.
- Se implementan los modelos en R.
- Se observan los resultados obtenidos.
- Se visualiza el modelo SVM obtenido.

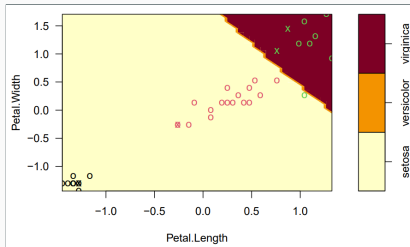
## Modelo SVM

```
# SVM with svm
model.svm = svm(Species ~ . ,data = train, method="C-classification", kernel="radial",
               gamma=0.1, cost=10)
# Predict
pred.svm = predict(model.svm,newdata = test)
```

## Resultados obtenidos

Modelo	Accuracy
Árboles de decisión	0.9736842 %
kNN	0.9473684 %
SVM	0.9736842 %
Random Forest	0.9736842 %

## Gráfica SVM

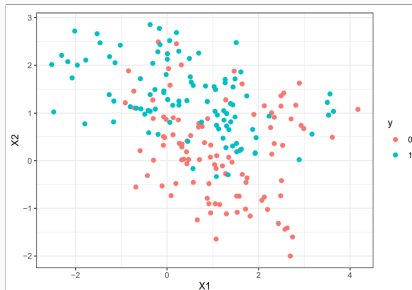


# Función lineal de 2 dimensiones

## Implementación

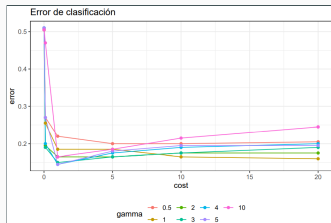
- Se obtienen los datos de *Elements of Statistical Learning*.
- *ESL.mixture* contiene valores de dos predictores etiquetados.
- Se generan modelos y se escoge el mejor.
- Se comprueba sobre nuevos elementos.

## Distribución de los datos

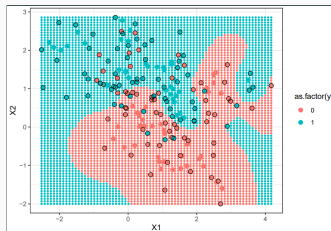


# Función lineal de 2 dimensiones

## Errores de clasificación



## Clasificación modelo





# Comparativa entre algoritmos

---

## Regresión logística

Ventajas	Desventajas
Útil para comprender la influencia de varias variables independientes en una sola variable de resultado.	Funciona solo cuando la variable predicha es binaria. Asume que todos los predictores son independientes entre sí. Asume que los datos no tienen valores perdidos.

## Naive Bayes

Ventajas	Desventajas
Requiere una pequeña cantidad de datos de entrenamiento para estimar los parámetros necesarios. Son extremadamente rápidos en comparación con métodos más sofisticados.	Existen algoritmos que clasifican mejor a pesar de tener un mayor tiempo de ejecución.

## Gradiente Descendente Estocástico

Ventajas	Desventajas
Eficiencia y facilidad de implementación.	Requiere una configuración de hiperparámetros. Sensible a normalización.

## K-Nearest Neighbors (kNN)

Ventajas	Desventajas
Simple de implementar. Robustez frente a ruido. Efectividad con datos de entrenamiento son grandes.	Requiere determinar k. Coste computacional alto.

## Árboles de decisión

Ventajas	Desventajas
Fácil de entender y visualizar. Requiere poca preparación de datos. Puede manejar tanto datos numéricos como categóricos.	Decidir grado de profundidad. Sensibles al ruido.

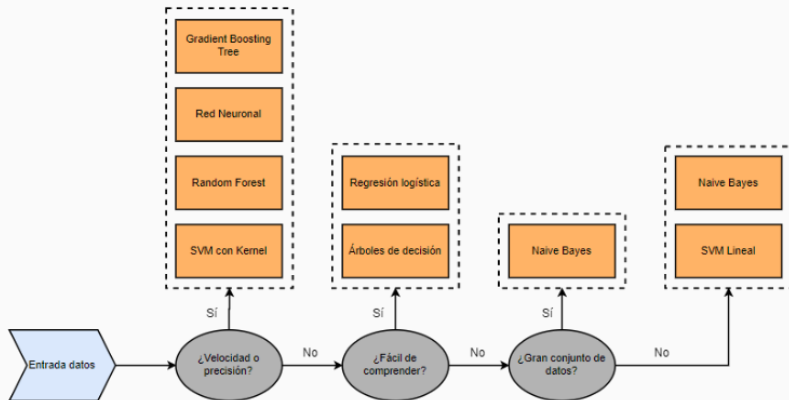
## Random Forest

Ventajas	Desventajas
Más avanzado que árboles de decisión en muchos casos. Reduce el sobreajuste.	Predicción lenta. Difícil de implementar. Alta complejidad. Gran gasto computacional.

## SVM

Ventajas	Desventajas
<p>Eficaz en espacios de alta dimensión.</p> <p>Eficiente en la memoria.</p> <p>Utiliza un subconjunto de puntos de entrenamiento en la función de decisión.</p>	<p>No proporciona estimaciones de probabilidad se calculan mediante una validación cruzada de cinco veces.</p>

## Elección del algoritmos según características



## Conclusiones

---

# Conclusiones

- SVM se centra en la maximización de los márgenes de separación entre diferentes clases mediante vectores de soporte con el fin de diferenciar entre clases en base a características espaciales.
- SVM da garantías cuando el conjunto de datos no es grande, hay un margen de separación y no mucho ruido o solapamiento.
- Numerosas aplicaciones de esta técnica como el diagnóstico de Parkinson y Temblor Esencial.
- SVM funciona en escenarios en las que prima la velocidad de ejecución y la precisión. Para ello se puede utilizar un kernel no lineal, mientras que en problemas sencillos SVM obtiene buenos resultados con kernel lineal con un conjunto de datos que no es excesivamente grande.



**Gracias por su atención**