

Support Vector Machine



Delta Analytics construye capacidad técnica alrededor del mundo.



El contenido de este curso está siendo desarrollado activamente por Delta Analytics, una organización sin fines de lucro 501(c)3 del Área de la Bahía que apunta a capacitar a las comunidades para aprovechar sus datos.

Por favor comuníquese con cualquier pregunta o comentario a inquiry@deltanalytics.org.

Descubre más sobre nuestra misión [aquí](#).



Support vector machine (SVM)

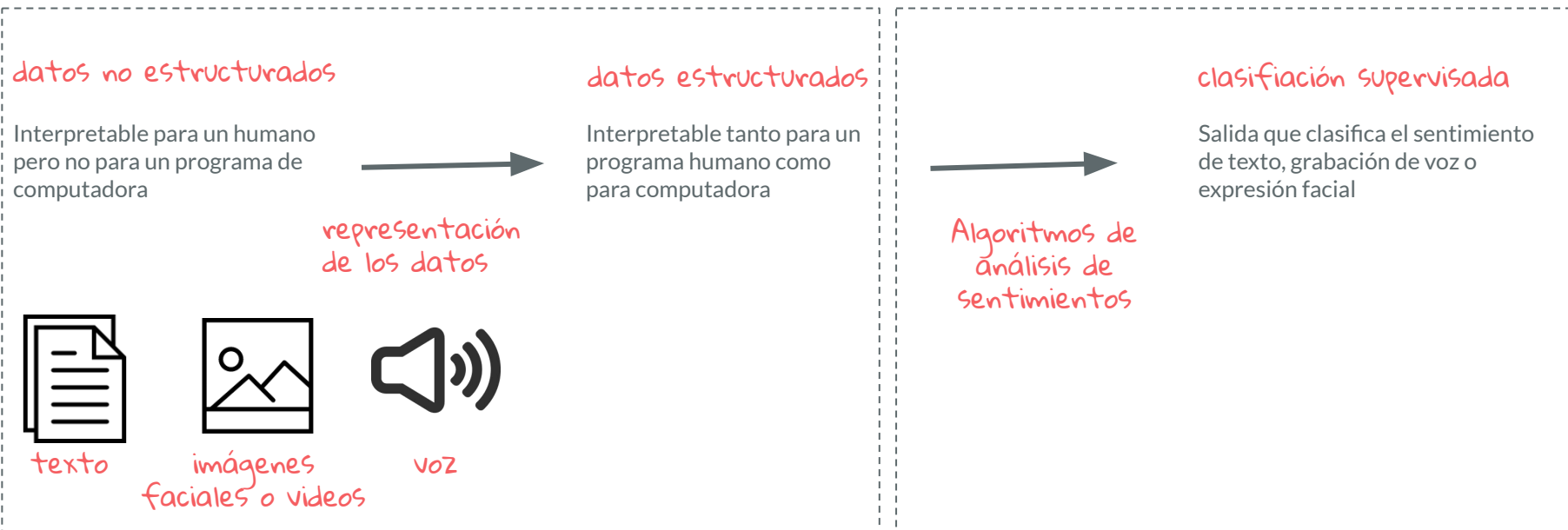


Checklist del módulo

- **Tarea**
 - ☐ Hiperplano (borde de desición)
 - ☐ Truco del kernel
- **Metodología de aprendizaje**
 - ☐ Optimización cuadrática
- **Medidas de desempeño**
 - ☐ Matriz de confusión
 - ☐ Gamma



El análisis de sentimiento usa datos estructurados para clasificar polaridad del texto

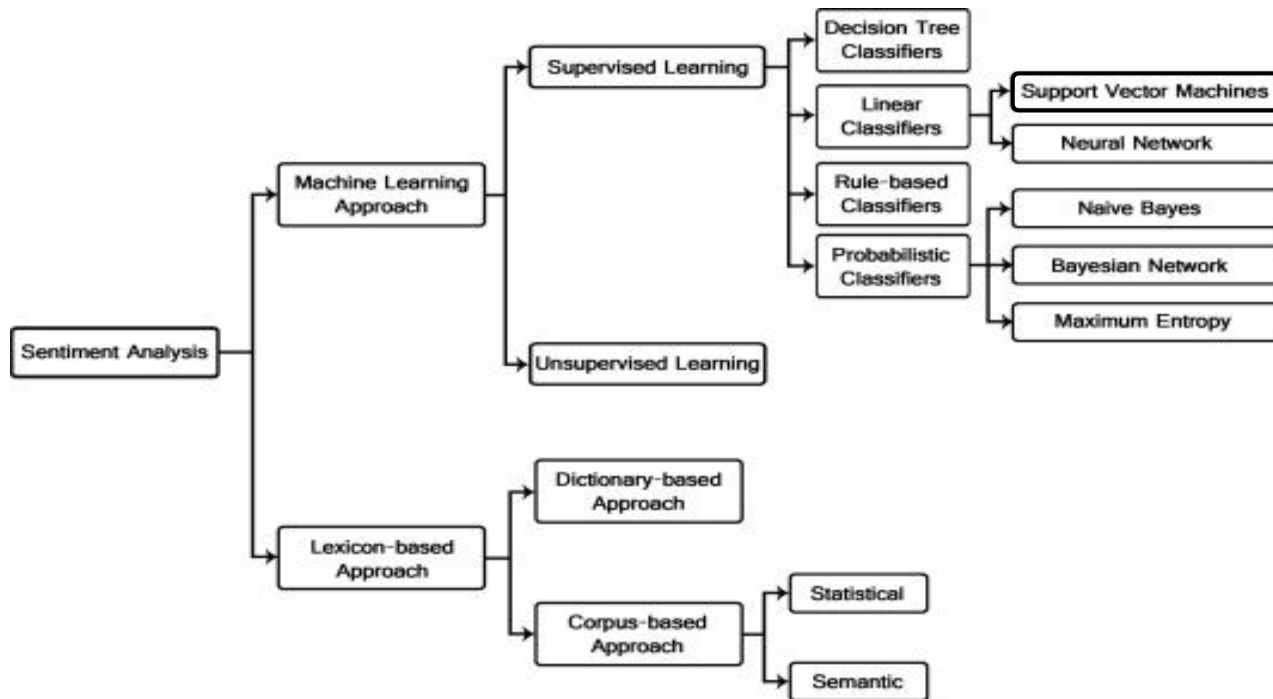


Bases del procesamiento del lenguaje natural.
Se verá en más detalle en un próximo módulo.

Análisis de sentimiento
foco de este módulo

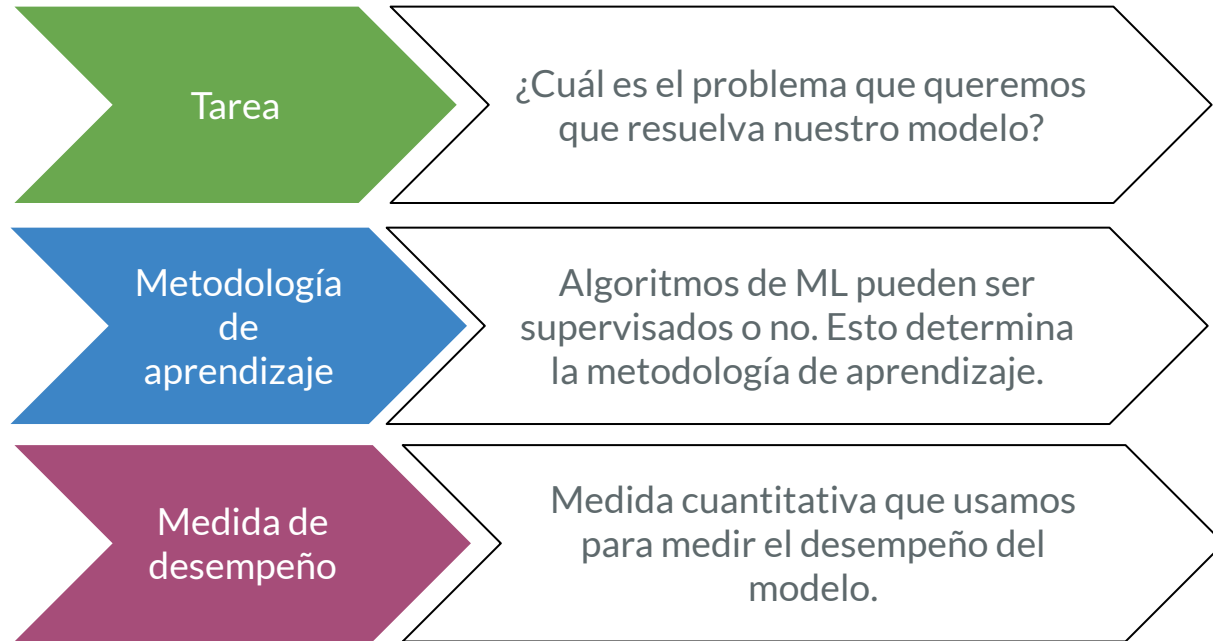


Aprenderemos como usar SVM para predecir el sentimiento del texto



Support vector machines son el foco de este módulo

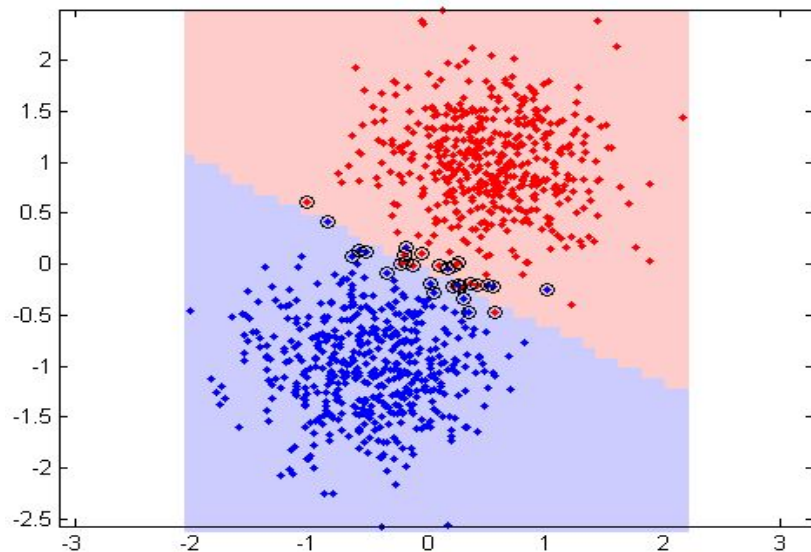
Usaremos nuestro ya familiar marco para introducir SVM



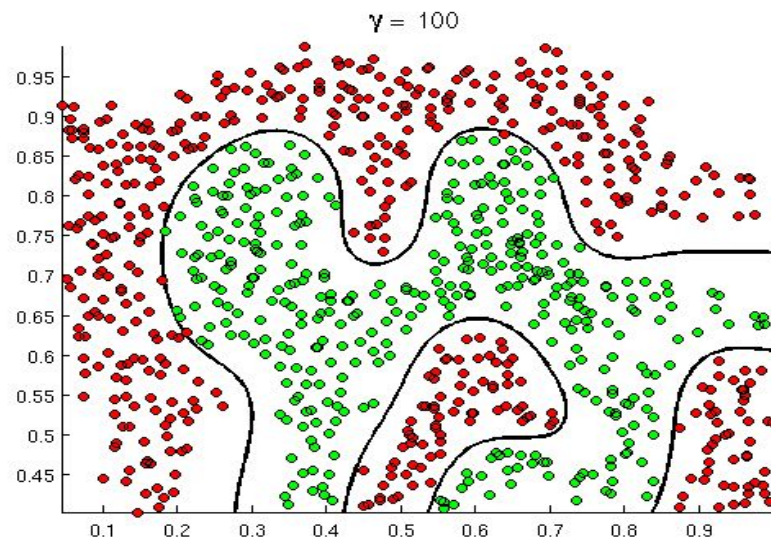
Tarea

¿Qué problema
queremos
resolver?

SVMs son una forma de clasificación que dibujan bordes para clasificar datos



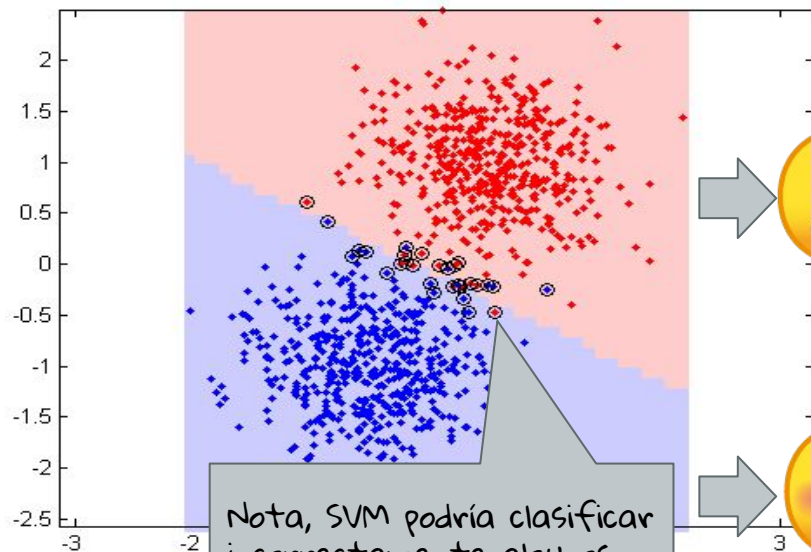
¿Cómo se relacionaría el análisis de
sentimiento con el gráfico de arriba?



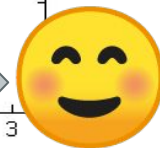
Tarea

¿Qué problema
queremos
resolver?

Áreas divididas por SVM definen las predicciones de sentimiento



Por ejemplo puntos de datos rojos representan textos clasificados como sentimiento negativo...



...mientras que puntos azules son de sentimiento positivo

En realidad, un gráfico representando análisis de sentimiento tendría más de dos dimensiones

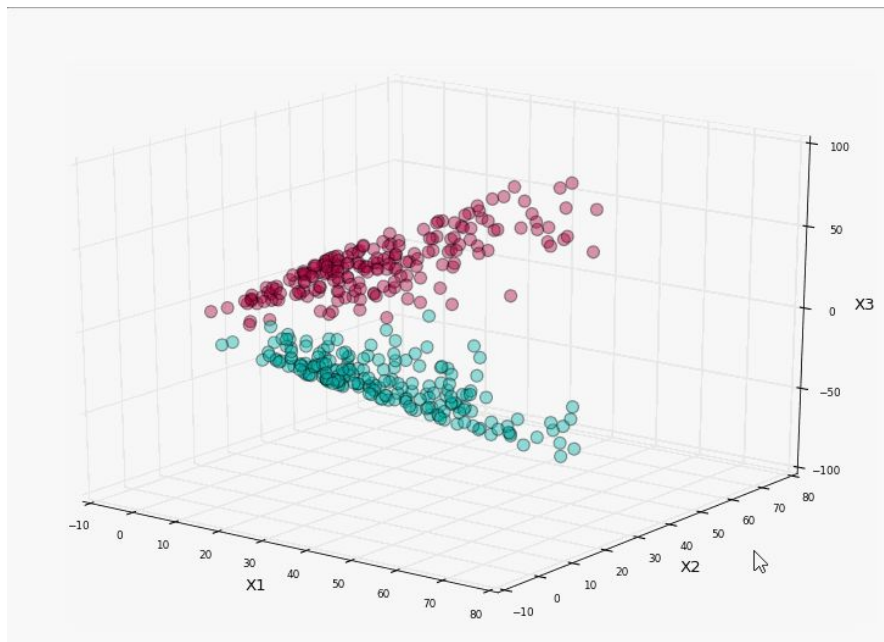
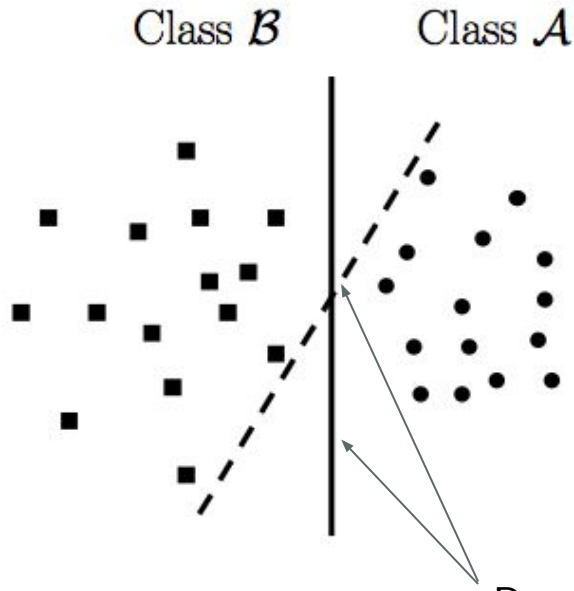


Gráfico	Sentiment analysis equivalent
Puntos de datos <ul style="list-style-type: none">• Rojo• Verde	Documento o texto a clasificar: <ul style="list-style-type: none">• Clasificado como negativo• Clasificado como positivo
Eje: x1	Frecuencia de palabra x1 en el texto
Eje: x2	Frecuencia de palabra x2 en el texto
Eje: x3	Frecuencia de palabra x3 en el texto
...	Cada palabra agrega otra dimensión (difícil de visualizar)

Tarea

Definiendo $f(x)$

La línea dividiendo los datos se llama borde de decisión o hiperplano



Dos ejemplos de
bordes de decisión

Como los árboles de decisión, un SVM representa puntos con valores a lo largo de varias características y un resultado de clasificación

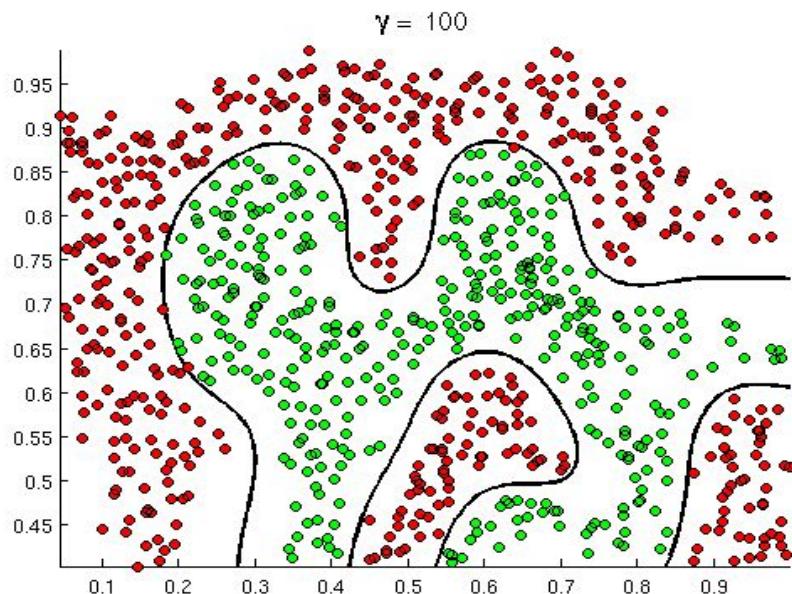
Nuestro $f(x)$ es el "hiperplano" óptimo dividiendo las clases

En el ejemplo a la izquierda, el hiperplano sólido es mejor borde que el segmentado.

Tarea

¿Es $f(x)$ el
correcto para
este
problema?

A veces el borde es no lineal sino que bastante curvo



Hasta ahora, en nuestros ejemplos de SVM los puntos pueden ser divididos por un plano

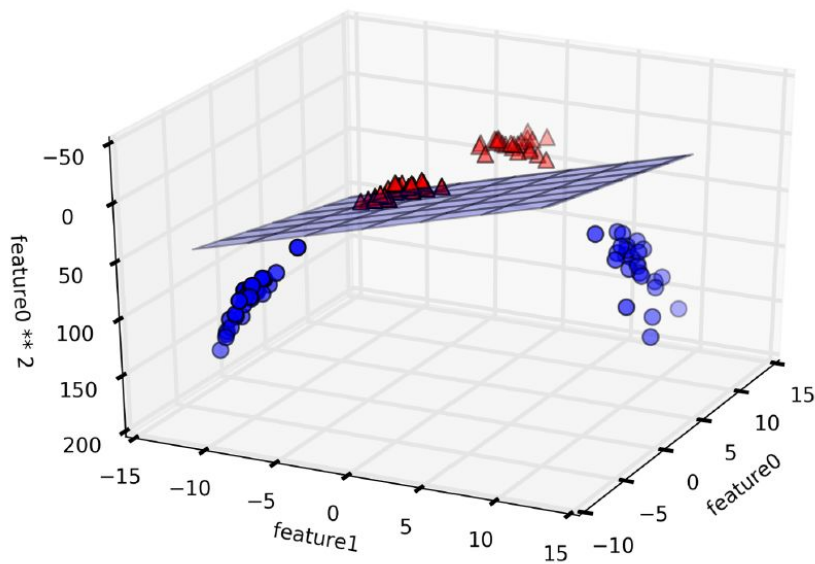
Estos datos se llaman linealmente separables

Sin embargo, ¿qué hay de los datasets que se parecen al de la izquierda?

Tarea

¿Es $f(x)$ el
correcto para
este
problema?

Para crear un hiperplano no lineal, podemos usar un truco matemático llamado kernels

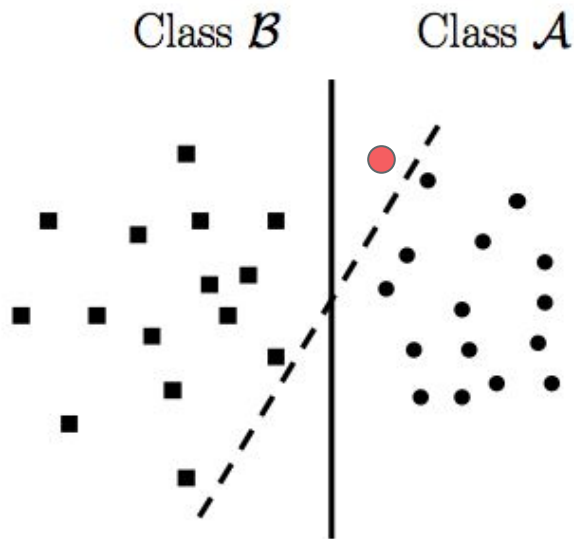


Teóricamente, si encontramos la combinación de características adecuada, podemos mapear los puntos a un espacio donde son linealmente separables

Esto es computacionalmente caro.

El "truco del kernel" es un atajo matemático que puede usar SVMs para crear bordes no lineales.

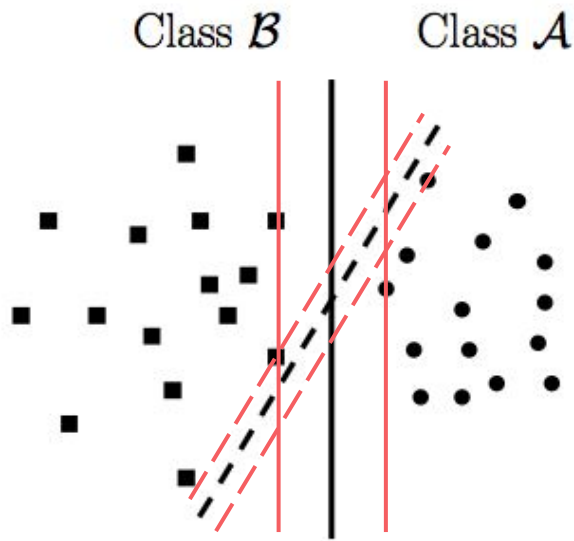
No todos los hiperplanos son iguales. Algunos son mejores que otros



Antes mencionamos que el hiperplano sólido era mejor que el segmentado. ¿Por qué?

Intuitivamente, el sólido divide los datos con mayor claridad.

Cuando agregamos el punto rojo de clase A, la línea segmentada lo hubiese clasificado erróneamente.



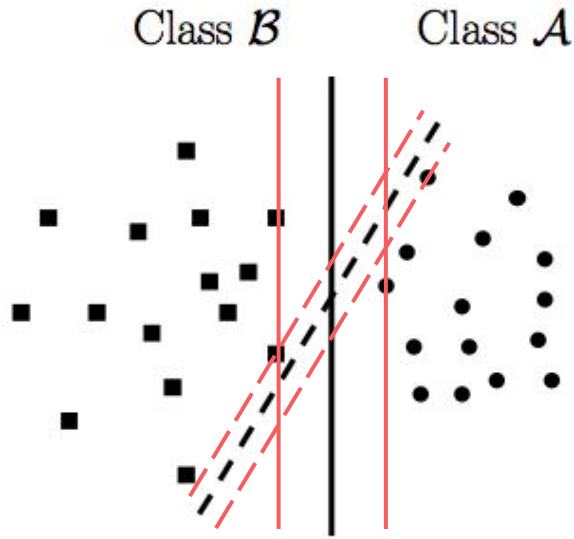
¿Qué son las líneas rojas que hemos dibujado?

Son líneas paralelas al borde de decisión que tocan un punto en cada clase

Los puntos que las líneas tocan se llaman vectores de soporte

¿Qué línea roja (sólida o segmentada) parece un mejor hiperplano para tí?

La línea sólida es óptima.



Notar como hay una mayor distancia entre las líneas rojas sólidas, que entre las líneas rojas segmentadas.

Si estás curioso, matemáticamente esto se puede expresar como un método de optimización cuadrática.

Minimiza la distancia
entre los bordes de
decisión

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{\|\mathbf{w}\|}{2}, \\ \text{s. t.} \quad & y^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + b) \geq 1 \quad \forall i \in \{1, \dots, N\} \end{aligned}$$

De modo que todos los
puntos son clasificados
correctamente como -1 or 1

Nota: La ecuación a la izquierda es una versión simplificada.

En realidad, SVMs permiten:

- Que algunos puntos sean mal clasificados
- Más de dos clases (no solo -1 y 1)

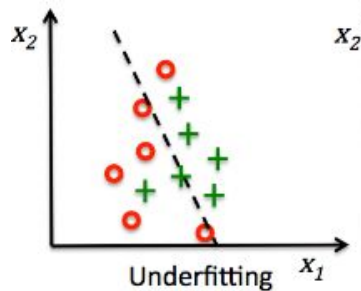
Desempeño

Medida cuantitativa para evaluar el desempeño del modelo.

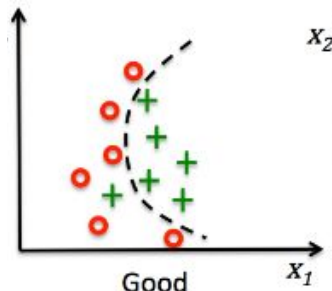


Como con todos los modelos, tenemos un tradeoff entre sesgo y varianza cuando evaluamos el desempeño

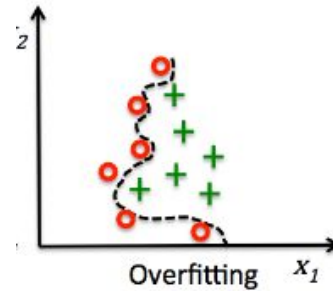
Alto sesgo



Good



Alta varianza



Desempeño

Medidas de
desempeño



Introduciré una herramienta común
para evaluar el desempeño de
modelos de clasificación

Clase predicha

Clase real

	Positiva	Negativa
Positiva	Verdadero positivo (Acertado)	Falso negativo
Negativa	Falso positivo (Error de tipo 1)	Verdadero negativo (Acertado)



Términos claves

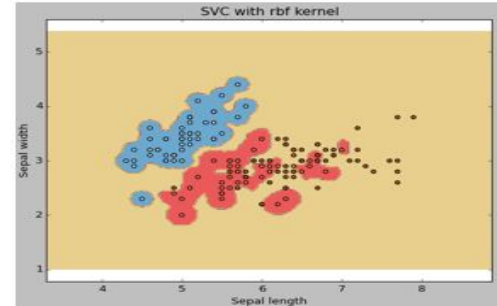
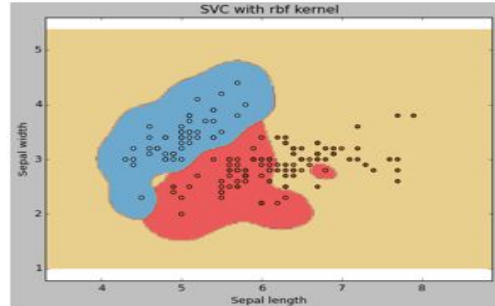
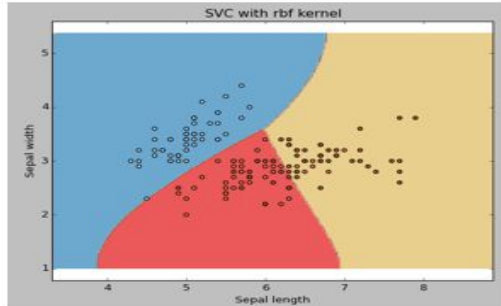
- Recall: Probabilidad condicional de verdaderos positivos dada la clase real
- Precisión: Probabilidad condicional de verdaderos positivos dada la clase predicha

		Clase predicha	
		Positiva	Negativa
Clase real	Positiva	Verdadero positivo (Acertado)	Falso negativo
	Negativa	Falso positivo (Error de tipo 1)	Verdadero negativo (Acertado)

Desempeño

Flexibilidad del
modelo

Puedes ajustar tus modelos de SVM con gamma, un input de sensibilidad. Entre más alto gamma, menor influencia tiene un punto individual en el modelo



¿Cómo se relaciona gama con el sesgo y la varianza?



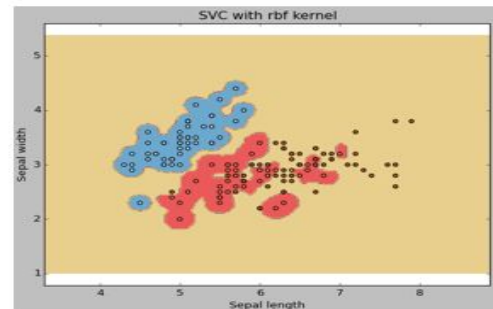
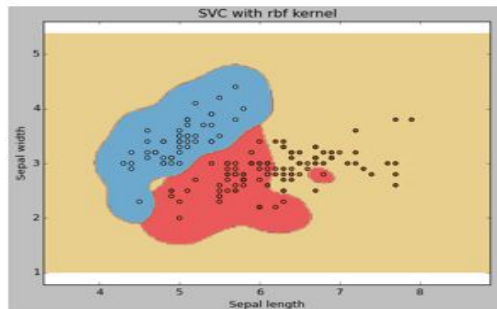
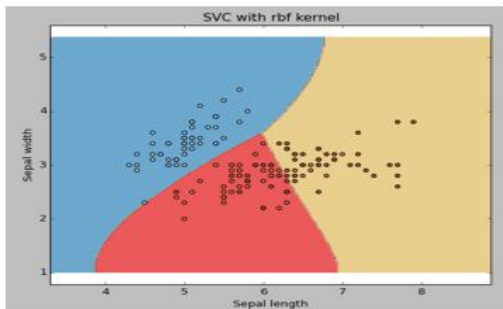
Desempeño

Flexibilidad del
modelo

Resulta que un alto gama puede dar lugar a sobre-ajuste, o alta varianza.

Bajo gamma

Alto gamma



¿Cómo se relaciona gama con el sesgo y la varianza?



modelo SVM:

Pros

- Puede ser usado con datos poco densos y desbalanceados
- Encuentra el borde óptimo de separación (comparado a otros clasificadores como perceptrones)
- Ocupa poca memoria para almacenar
- Fácil de agregar características dado el “truco del kernel”

Cons

- Los datos deben ser linealmente separables para “hard-margin” SVMs
- Se usan supuestos adicionales cuando los datos no son linealmente separables
- Provee clasificación determinista (no estimado probabilístico)



Recursos teóricos que recomendamos:

Support vector machines

- Introduction to Machine Learning with Python (O'Reilly Media) Chapter 2, Linear Models
- [An Idiot's Guide to Support Vector Machines \(R. Berwick\)](#)
- [Sentiment analysis using Support Vector Machine \(Nurulhuda Zainuddin\)](#)



Felicidades! ¡Terminaste el módulo!

Obtén más información sobre el machine learning de Delta para una buena misión aquí.