

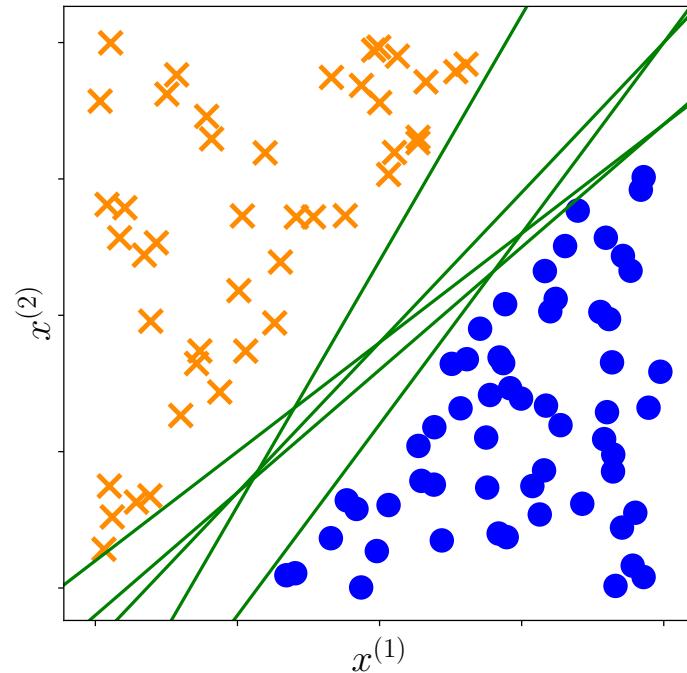
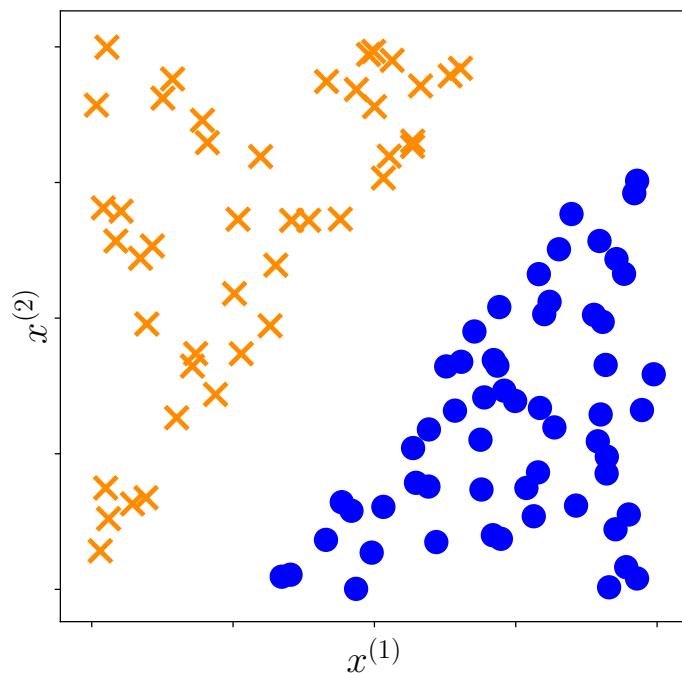
UNIVERSIDAD
PANAMERICANA

Máquinas de Soporte Vectorial

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un *margen* y un *kernel*.

Asumiendo un conjunto de datos de dos clases $\{+1, -1\}$ linealmente separables, SVM se comporta como un clasificador binario que divide las clases mediante un hiperplano:

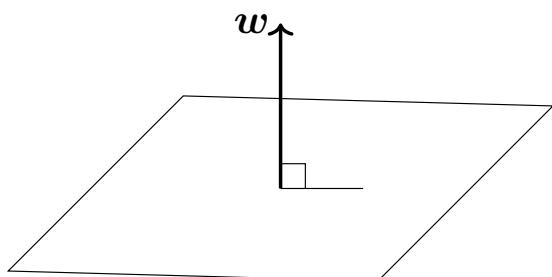


Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

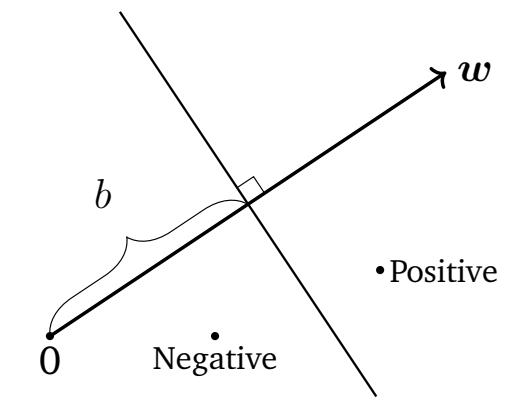
Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

Asumiendo un conjunto de datos de dos clases $\{+1, -1\}$ linealmente separables, SVM se comporta como un clasificador binario que divide las clases mediante un **hiperplano**:



(a) Separating hyperplane in 3D



(b) Projection of the setting in (a) onto a plane

Hiperplano:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

$$f(\mathbf{x}) = 0$$

Con producto punto:

$$f(\mathbf{x}) = w_1 x_1 + \cdots + w_D x_D + b$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

Asumiendo un conjunto de datos de dos clases $\{+1, -1\}$ linealmente separables, SVM se comporta como un clasificador binario que divide las clases mediante un **hiperplano**:

Hiperplano:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

$$f(\mathbf{x}) = 0$$

$$f(\mathbf{x}_a) \geq 0, \Rightarrow y_a = +1$$

$$f(\mathbf{x}_a) < 0, \Rightarrow y_a = -1$$

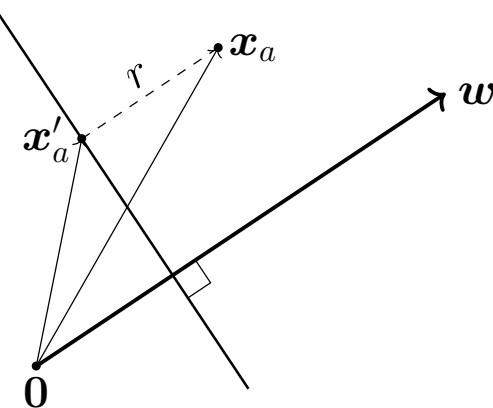
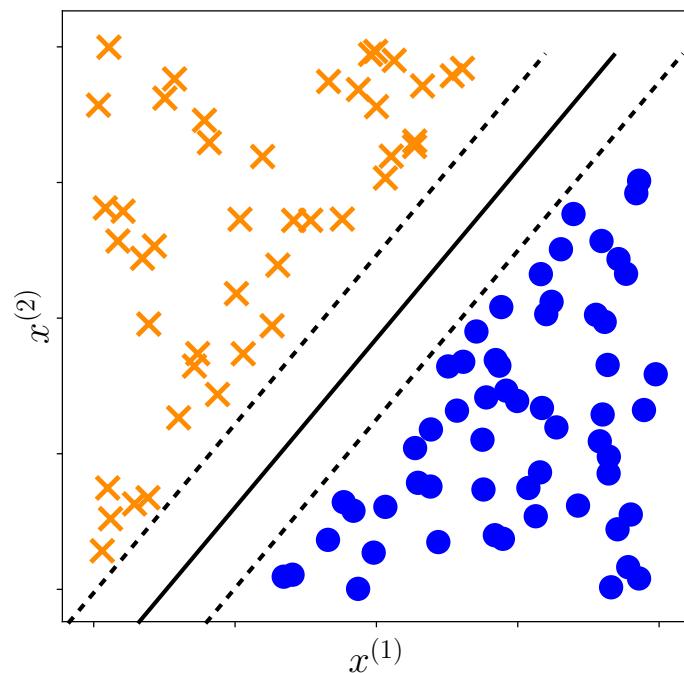
$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 0$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margin** y un **kernel**.

El hiperplano debe tener un **margin** que garantice que los datos están correctamente separados. Por lo que se buscará un hiperplano con el máximo margen posible:

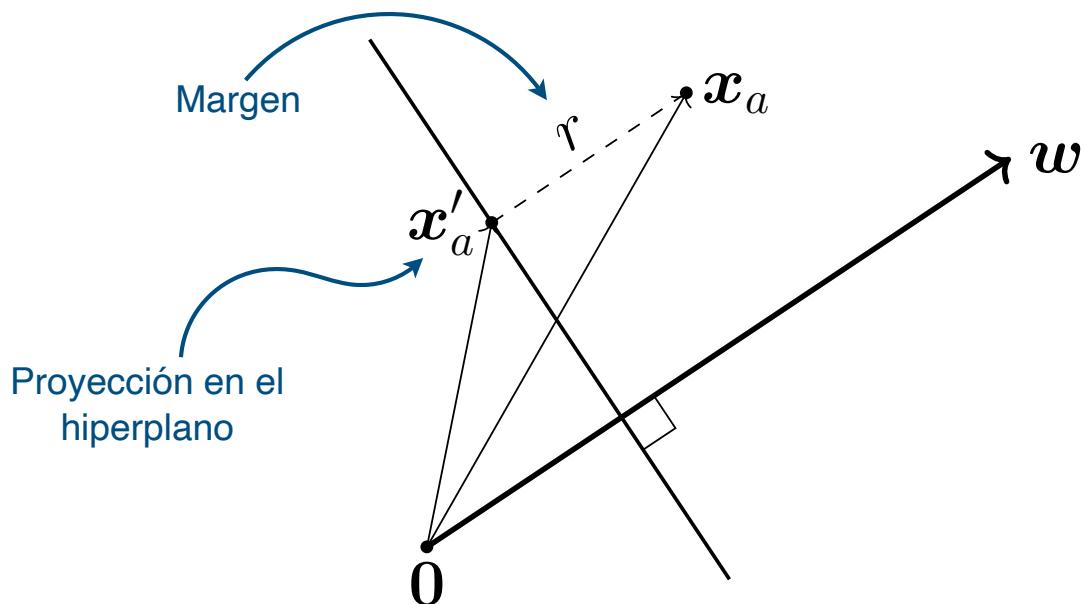


Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margin** y un **kernel**.

El hiperplano debe tener un **margin** que garantice que los datos están correctamente separados. Por lo que se buscará un hiperplano con el máximo margen posible:



$$\mathbf{x}_a = \mathbf{x}'_a + r \frac{\mathbf{w}}{\|\mathbf{w}\|}$$

Condición:

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq r$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en la separación lineal de las clases con un **margin** y un **kernel**.

El hiperplano debe tener un **margin** que garantice que los datos están correctamente separados. Por lo que se buscará un hiperplano con el máximo margen posible:

$$\max_{\mathbf{w}, b, r} r$$

$$r = \frac{1}{\|\mathbf{w}\|}$$

Sujeto a: $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq r$

$$\|\mathbf{w}\| = 1$$

$$r > 0$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

El **margen duro de un SVM** se define como la solución del proceso de optimización:

$$\max_{\mathbf{w}, b} \frac{1}{\|\mathbf{w}\|}$$

Sujeto a: $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1$

De manera equivalente:

$$\max_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$

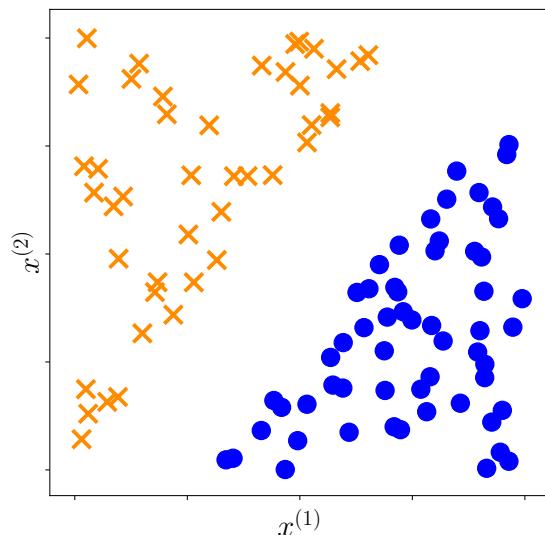
Sujeto a: $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

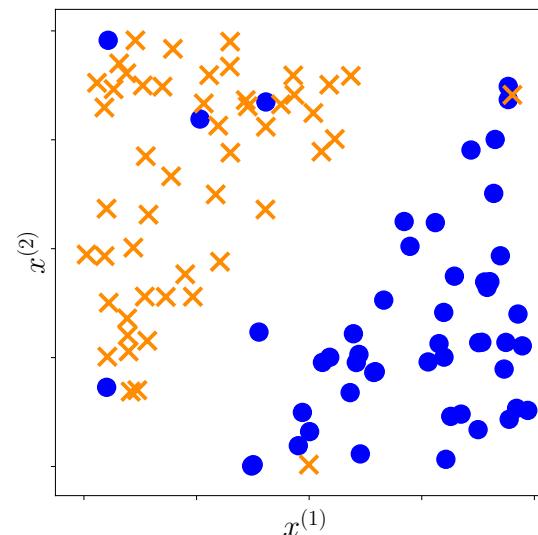
Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

El **margen suave de un SVM** permite relajar el problema de optimización para permitir el uso de SVM en conjuntos de datos no separables linealmente:



(a) Linearly separable data, with a large margin



(b) Non-linearly separable data

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

El **margen suave de un SVM** permite relajar el problema de optimización para permitir el uso de SVM en conjuntos de datos no separables linealmente:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \max\{0, 1 - y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)\}$$

Regularización

Función de pérdida

$$C > 0$$

Parámetro de regulación
del margen

Pérdida de la bisagra (*hinge loss*):

$$\ell(t) = \max\{0, 1 - t\}$$

$$t = y(\langle \mathbf{w}, \mathbf{x} \rangle + b)$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en la separación lineal de las clases con un **margin** y un **kernel**.

Un problema en el enfoque anterior es la dependencia del número de parámetros respecto al número de atributos; por lo que se propone un problema de optimización equivalente:

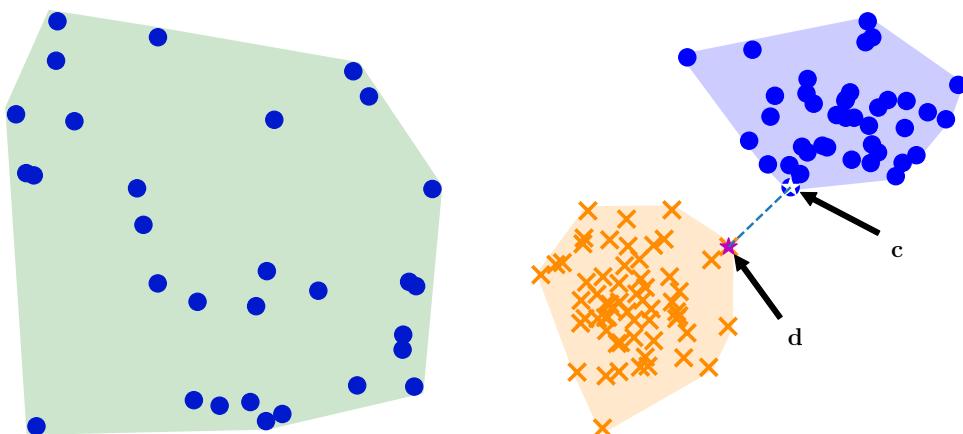
- * **SVM primitivo** – problema de optimización donde el número de parámetros depende del número de atributos. Crece rápidamente y el proceso computacional se vuelve pesado.
- * **SVM dual** – problema de optimización donde el número de parámetros depende del número de muestras y permite la introducción de **kernels**.

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

El problema de optimización SVM dual se puede interpretar como encontrar la **envolvente convexa** más pequeña para cada clase:



(a) Convex hull.

(b) Convex hulls around positive (blue) and negative (orange) examples. The distance between the two convex sets is the length of the difference vector $c - d$.

Descripción de una envolvente convexa:

$$conv(\mathbf{X}) = \left\{ \sum_i \alpha_i \mathbf{x}_i \right\}$$

$$\sum_i \alpha_i = 1$$

$$\alpha_i \geq 0$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

El problema de optimización SVM dual se puede interpretar como encontrar la **envolvente convexa** más pequeña para cada clase:

$$\min_{\alpha} \frac{1}{2} \left\| \sum_{i:y_i=+1} \alpha_i^+ \mathbf{x}_i - \sum_{i:y_i=-1} \alpha_i^- \mathbf{x}_i \right\|^2$$

Sujeto a: $\sum_i y_i \alpha_i = 0$

$$\min_{\alpha} \frac{1}{2} \sum_i \sum_j y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_i \alpha_i$$

De manera equivalente:

Sujeto a: $\sum_i y_i \alpha_i = 0$

$$0 \leq \alpha_i \leq C$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margen** y un **kernel**.

SVM ocupa un hiperplano para hacer la separación de clases. Sin embargo, un problema con clases no separables linealmente no podrá ser resuelto de forma correcta. Entonces, se utiliza el **truco del kernel** en SVM dual para tratar clases no separables:

$$\min_{\alpha} \frac{1}{2} \sum_i \sum_j y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_i \alpha_i$$

Sujeto a: $\sum_i y_i \alpha_i = 0$

$$0 \leq \alpha_i \leq C$$

SVM dual

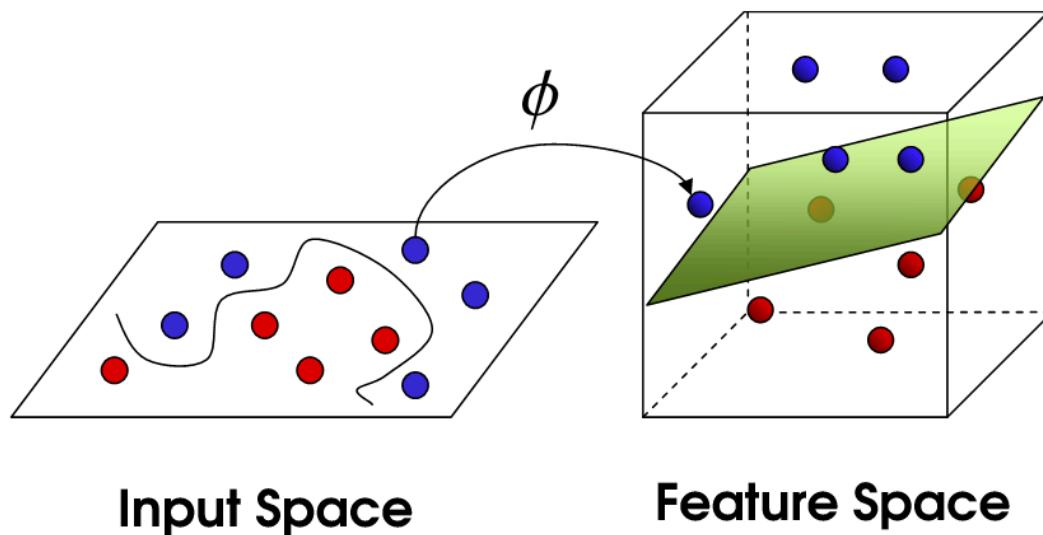
El producto interno se puede sustituir por una **función kernel**, o simplemente **kernel**:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Una función ***kernel*** toma dos vectores de entradas del espacio original y regresa el producto interno de los vectores en el espacio de características. El *kernel* es una función de similitud.



$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

Máquinas de soporte vectorial

Una **función *kernel*** toma dos vectores de entradas del espacio original y regresa el producto interno de los vectores en el espacio de características. El *kernel* es una función de similitud.

Funciones *kernel* comunes en SVM:

Lineal:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j + c$$

Polinomial:

$$k(\mathbf{x}_i, \mathbf{x}_j) = (a\mathbf{x}_i^T \mathbf{x}_j + c)^d$$

Gaussiano, o de base radial (RBF):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

Wilimitis, D., The Kernel Trick in Support Vector Classification, Towards Data Science, 2020.

Máquinas de soporte vectorial

Una **función *kernel*** toma dos vectores de entradas del espacio original y regresa el producto interno de los vectores en el espacio de características. El *kernel* es una función de similitud.

Un ejemplo de cómo usar la función *kernel* (polinomial):

Mapeo de 2D a 3D:

$$\Phi(\mathbf{x}) \rightarrow x_1^2, x_2^2, \sqrt{2}x_1x_2$$

Usando el mapeo directo:

$$\begin{aligned}\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle &= \langle \{x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\}, \{x_{j1}^2, x_{j2}^2, \sqrt{2}x_{j1}x_{j2}\} \rangle \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

Función *kernel*:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle^2$$

Usando el *kernel*:

$$\begin{aligned}\langle \mathbf{x}_i, \mathbf{x}_j \rangle^2 &= \langle \{x_{i1}, x_{i2}\}, \{x_{j1}, x_{j2}\} \rangle^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

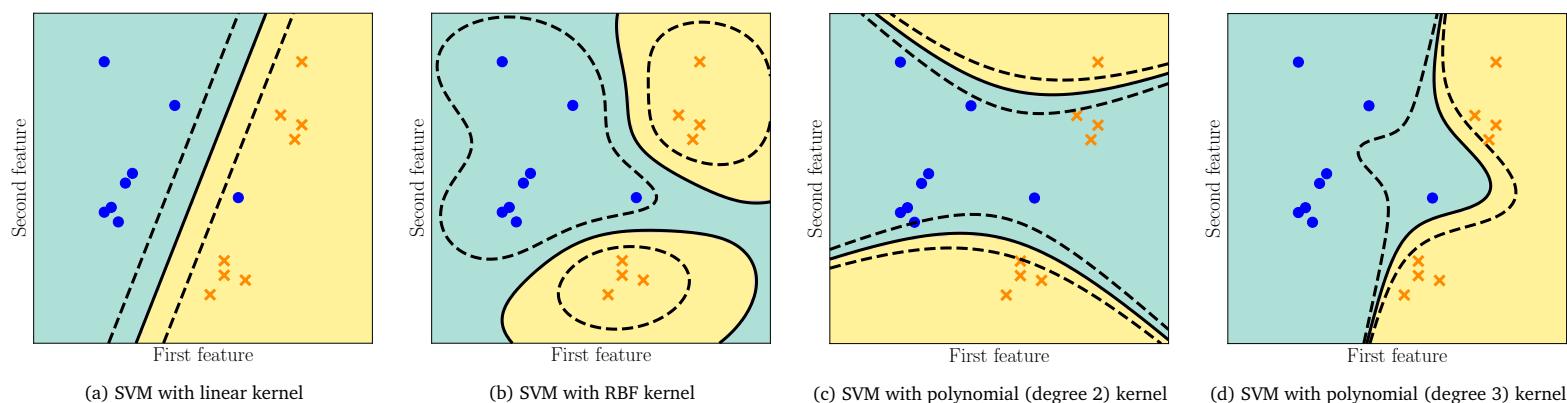
Wilimitis, D., The Kernel Trick in Support Vector Classification, Towards Data Science, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un *margen* y un *kernel*.

Al proceso de utilizar una función *kernel* en SVM dual se le conoce como el **truco del kernel**. Toma este nombre porque no es necesario tener de forma explícita el mapa al espacio de características; lo que ahorra procesamiento computacional.

El entrenamiento de un SVM se reduce a encontrar los coeficientes duales y los parámetros del *kernel*.

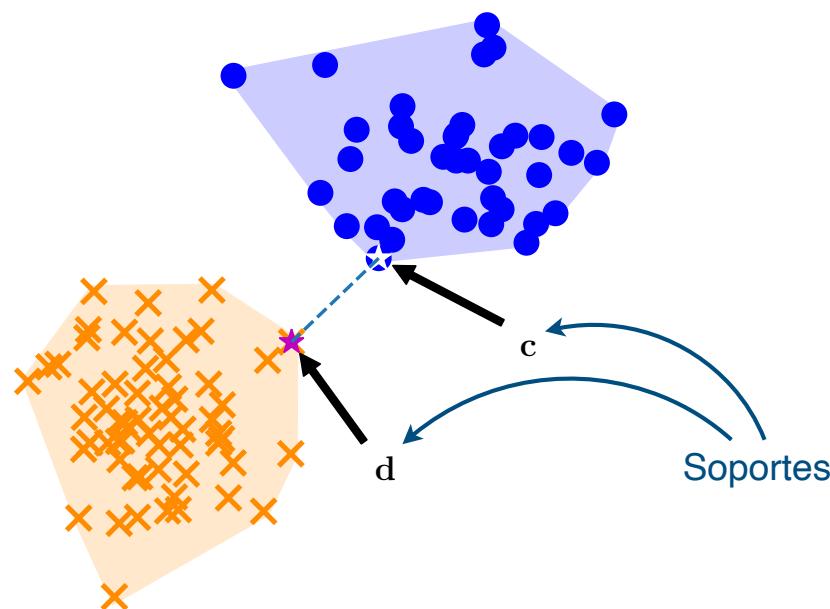


Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.

Máquinas de soporte vectorial

Las **máquinas de soporte vectorial (SVM)**, por sus siglas en inglés) es un modelo de aprendizaje supervisado que permite hacer clasificación basada en las separación lineal de las clases con un **margin** y un **kernel**.

Los **vectores de soporte** son aquellos ejemplos del conjunto de datos que sirven para definir el margen del hiperplano.



Deisenroth, M., Faisal, A., Ong, C., Mathematics for Machine Learning, Cambridge University Press, 2020.