

Support Vector Machine

SVM

Dr. Mauricio Toledo-Acosta

Diplomado Ciencia de Datos con Python

Table of Contents

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- 1 Introducción
- 2 SVM de margen duro
- 3 SVM de margen suave
- 4 Kernel Trick
- 5 Métricas de desempeño

Introducción

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

Introducción

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

Support Vector Machine

Modelo supervisado de clasificación binaria que busca encontrar una frontera de decisión óptima que separe las clases de puntos.

El modelo lineal de clasificación

Los puntos x que satisfacen $y(x) = w^T \cdot x + w_0 = 0$ forman la frontera de decisión (FD), la cual divide al espacio de datos en dos regiones.

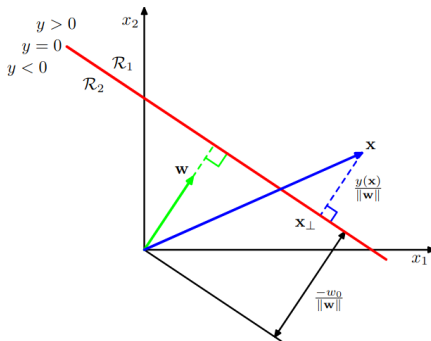


Table of Contents

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- 1 Introducción
- 2 SVM de margen duro
- 3 SVM de margen suave
- 4 Kernel Trick
- 5 Métricas de desempeño

SVM de margen duro

Support
Vector
Machine

Clasificación

Introducción

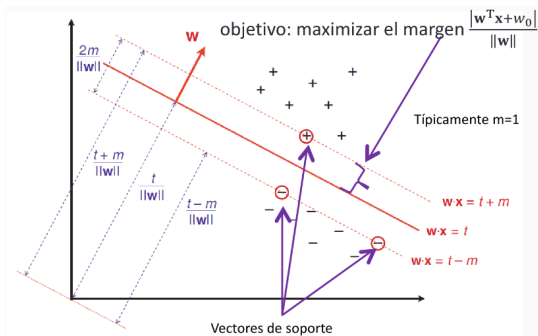
SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

Analicemos el caso con datos linealmente separables. Cambiaremos ligeramente la notación, FD está definida por $g(x) = w^T \cdot x - t = 0$. Queremos encontrar una FD con margen $m = 1$.



Planteamiento del problema

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

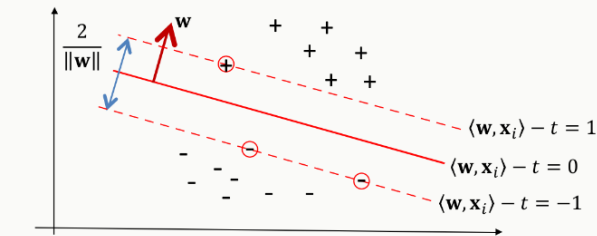
Métricas de
desempeño

► Nuestro objetivo es:

$$\mathbf{w}^*, t^* = \operatorname{argmin}_{\mathbf{w}, t} \frac{1}{2} \|\mathbf{w}\|^2$$

► Sujeto a las siguientes N restricciones:

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - t) \geq 1, \quad 1 \leq i \leq N$$



Planteamiento del problema

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

► Definimos el lagrangiano

$$\begin{aligned}\mathcal{L}_P(\mathbf{w}, t, \alpha_1, \dots, \alpha_N) &= \overset{\text{Minimizar}}{\downarrow} \frac{1}{2} \|\mathbf{w}\|^2 - \overset{\text{Maximizar}}{\downarrow} \sum_{i=1}^N \alpha_i (y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle - t) - 1) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i y_i \langle \mathbf{w}, \mathbf{x}_i \rangle + \sum_{i=1}^N \alpha_i y_i t + \sum_{i=1}^N \alpha_i \\ &= \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle - \langle \mathbf{w}, \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \rangle + t \left(\sum_{i=1}^N \alpha_i y_i \right) + \sum_{i=1}^N \alpha_i\end{aligned}$$

► Para un t óptimo $\partial_t \mathcal{L}_P = 0 \Rightarrow \sum_{i=1}^N \alpha_i y_i = 0$

► Para pesos óptimos $\partial_{\mathbf{w}} \mathcal{L}_P = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$

Planteamiento del problema

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- Reinsertando estas expresiones en \mathcal{L}_P obtenemos \mathcal{L}_D el lagrangiano del problema dual:

$$\mathcal{L}_D(\alpha_1, \dots, \alpha_N) = -\frac{1}{2} \left\langle \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i, \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \right\rangle + \sum_{i=1}^N \alpha_i$$

$$= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^N \alpha_i$$

Planteamiento del problema

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- *El problema de optimización dual es el siguiente:*

$$\alpha_1^*, \dots, \alpha_N^* = \operatorname{argmax}_{\alpha_1, \dots, \alpha_N} -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^N \alpha_i$$

- *Sujeto a las restricciones:*

$$\alpha_i > 0, \quad 1 \leq i \leq N \quad \text{y} \quad \sum_{i=1}^N \alpha_i y_i = 0$$

Ejemplo

Support
Vector
Machine

Clasificación

Introducción

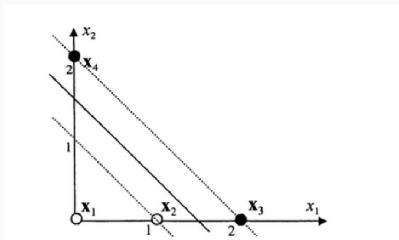
SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- Encuentra W óptimo para este problema: $X_1=[0,0]$ $X_2=[1,0]$ para la clase (+1) y $X_3=[2,0]$ y $X_4=[0,2]$ para la clase (-1)



Ejemplo

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

$$\mathcal{L}_D(\alpha_1, \dots, \alpha_N) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$\mathcal{L}_D = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) - \frac{1}{2} (\alpha_2^2 - 4\alpha_2\alpha_3 + 4\alpha_3^2 + 4\alpha_4^2)$$

Diferenciando con respecto a los α 's y utilizando la restricción $\sum_{i=1}^N \alpha_i y_i = 0$ obtenemos:

$$\begin{cases} \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0 \\ \alpha_2 - 2\alpha_3 = 1 \\ -2\alpha_2 + 4\alpha_3 = 1 \\ 4\alpha_4 = 1 \end{cases}$$

de donde: $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = \frac{3}{4}$, $\alpha_4 = 1/4$

Aplicando: $\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$, finalmente obtenemos

$$\mathbf{w} = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}, w_0 = 3/4 \text{ y } d(x) = 3 - 2x_1 - 2x_2 = 0$$

Table of Contents

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- 1 Introducción
- 2 SVM de margen duro
- 3 SVM de margen suave
- 4 Kernel Trick
- 5 Métricas de desempeño

SVM de margen suave

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

$$\mathbf{w}^*, t^*, \xi_i^* = \underset{\mathbf{w}, t, \xi_i}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

sujeto a $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle - t) \geq 1 - \xi_i$ y $\xi_i \geq 0, 1 \leq i \leq N$

- C es un parámetro definido por el usuario que balancea la maximización del margen contra la minimización de las variables de holgura:
 - un valor alto de C significa que los errores de margen son altamente costosos,
 - un valor pequeño de C permite más errores de margen con tal de hacer mas grande el margen.
- Si permitimos más errores de margen necesitamos menos vectores de soporte, por lo tanto C controla la 'complejidad' de la SVM y por ello se le denomina el *parámetro de complejidad*.

SVM de margen suave

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- Buscamos soluciones mediante el nuevo Lagrangiano:

$$\begin{aligned}\mathcal{L}(\mathbf{w}, t, \xi_i, \alpha_i, \beta_i) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i (y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle - t) - (1 - \xi_i)) - \sum_{i=1}^N \beta_i \xi_i \\ &= \mathcal{L}(\mathbf{w}, t, \alpha_i) + \sum_{i=1}^N (C - \alpha_i - \beta_i) \xi_i\end{aligned}$$

- La solución óptima es tal que $\partial_{\xi_i} \mathcal{L} = 0 \Rightarrow$ el término añadido desaparece en el problema dual.
- Además, puesto que α_i y β_i son positivos, α_i no puede ser mayor a C :

$$\alpha_1^*, \dots, \alpha_N^* = \operatorname{argmax}_{\alpha_1, \dots, \alpha_N} -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^N \alpha_i$$

Sujeto a las restricciones: $0 \leq \alpha_i \leq C$, $1 \leq i \leq N$ y $\sum_{i=1}^N \alpha_i y_i = 0$

Table of Contents

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- 1 Introducción
- 2 SVM de margen duro
- 3 SVM de margen suave
- 4 Kernel Trick**
- 5 Métricas de desempeño

El truco del Kernel

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño



El truco del Kernel

Support
Vector
Machine

Clasificación

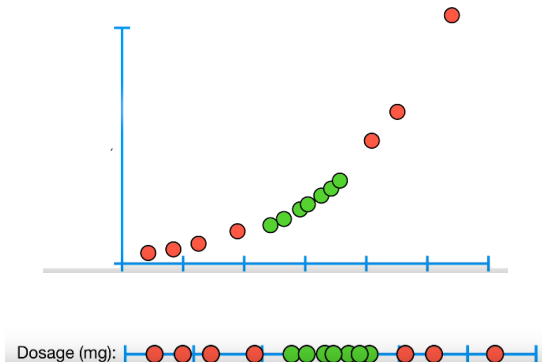
Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño



Tipos de Kernel

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- Aunque nuevos kernels aparecen en la literatura, los siguientes cuatro son básicos y ampliamente utilizados:

Lineal:	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$
Polinomial:	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = (\langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)^p, r \geq 0$
Gaussiano (<i>Radial Basis Function</i> – RBF):	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right) = \exp\left(-\gamma\ \mathbf{x}_i - \mathbf{x}_j\ ^2\right)$
Sigmoide:	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)$

donde r, p, γ son parámetros de los modelos.

Table of Contents

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

- 1 Introducción
- 2 SVM de margen duro
- 3 SVM de margen suave
- 4 Kernel Trick
- 5 Métricas de desempeño

Matriz de Confusión

Support
Vector
Machine

Clasificación

Introducción

SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Total population = P + N	Positive (PP)	Negative (PN)
	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

Métricas de desempeño

- **Accuracy:** De todos la población, ¿cuántos predije correctamente?

$$A = \frac{TP + TN}{\text{Total}}.$$

- **Recall:** De todos la población positiva, ¿cuántos predije correctamente como positivos?

$$R = \frac{TP}{TP + FN}.$$

- **Precision:** De todos los que predije como positivos, ¿cuántos son realmente positivos?

$$P = \frac{TP}{TP + FP}.$$

- **F1 score:** Media armónica de la precisión y el recall:

$$2 \frac{P \cdot R}{P + R}$$

Ejemplo

Support
Vector
Machine

Clasificación

Introducción


SVM de
margen duro

SVM de
margen suave

Kernel Trick

Métricas de
desempeño

Tenemos la siguiente población $\{+ + - - - -\}$:



+	+	-	-	-	-
-	-	-	-	-	-

Accuracy: 0.66, Recall: 0, Precision: 0.



+	+	-	-	-	-
-	-	-	-	-	-

Accuracy: 0.33, Recall: 1, Precision: 0.33.