
Máquinas de Vectores Soporte (SVM)

García Prado, Sergio
sergio@garciparedes.me

29 de mayo de 2017

Resumen

En este documento se realiza una descripción acerca del algoritmo de aprendizaje automático basado en máquinas de vectores soporte (SVM). Además, se resuelve de manera gráfica la tarea de aprendizaje mediante SVM sobre un conjunto de datos sencillo. Por último, se realizan distintos experimentos utilizando el conjunto de datos Wine para tratar de encontrar la función kernel óptima sobre dicho conjunto de datos.

1. INTRODUCCIÓN

1.1. MÁQUINAS DE VECTORES SOPORTE

Las máquinas de vector soporte son una estrategia de aprendizaje similar a la regresión lineal. Sin embargo, añade la restricción de que el hiperplano de separación entre los conjuntos de datos a clasificar debe poseer la propiedad de ser el de mayor distancia entre los dos conjuntos de datos.

De esta manera, el clasificador obtenido será mucho más preciso ya que la solución generaliza mejor la diferencia entre clases por lo que se reduce en mayor medida el sobreajuste. El algoritmo funciona mediante el apoyo entre los puntos de clases distintas más cercanos entre si, a los cuales se denomina vectores soporte. La razón de ello es que si el plano se construye tratando de maximizar la distancia a estos puntos, el resultado es equivalente a si se quisiera maximizar sobre el conjunto de datos completo.

2. EXPERIMENTO SOBRE CONJUNTO DE DATOS SENCILLO

En esta sección se describe de manera teórica cómo funcionaria la fase de aprendizaje sobre el conjunto de datos sencillo, el cual se muestra en la tabla 1, utilizando el algoritmo de *máquinas de vector soporte (SVM)*. Puesto que este algoritmo trata de maximizar la distancia del hiperplano de separación entre las dos clases teniendo en cuenta únicamente los puntos de clases distintas más cercanos entre sí los cuales se denominan vectores soporte, el resto de instancias no se tienen en cuenta para la construcción del hiperplano.

Para elegir el hiperplano óptimo por tanto hay que tratar de maximizar la distancia del mismo a las dos clases a clasificar. Puesto que en este caso se ha utilizado un kernel lineal, el hiperplano se puede representar como una recta sobre el espacio de dos dimensiones al cual pertenece el conjunto de datos. Para la construcción del mismo el

Simple Dataset		
x_1	x_2	$d(x)$
1	1	1
2	1	1
2	2	1
1	3	1
-1	1	-1
0	2	-1
-1	3	-1

Tabla 1: Conjunto de datos Simple

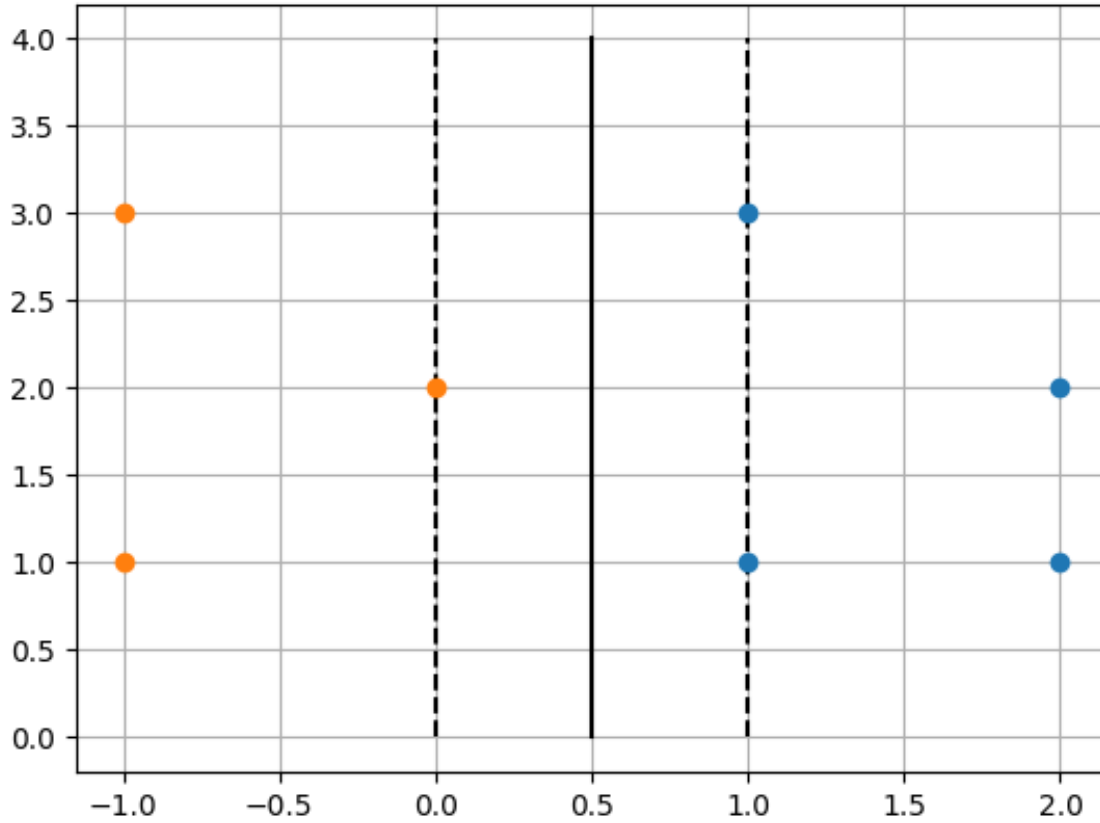


Figura 1: Representación gráfica del hiperplano construido a partir de máquinas de vector soporte SVM sobre el conjunto de datos simple

algoritmo de máquinas de vector soporte se apoya en la utilización de dos rectas paralelas entre si en las cuales se apoyan los vectores soporte de cada conjunto de datos. El hiperplano a construir por tanto, se debe encontrar en el punto intermedio de estas, por lo que también es paralelo a ellas.

Esta construcción se puede apreciar en la figura 1 de manera visual. Tal y como se puede ver, no existe un otro hiperplano para el conjunto de datos utilizado que maximice en mayor medida la distancia entre el hiperplano y los vectores soporte de las dos clases.

3. EXPERIMENTO SOBRE WINE DATASET

El conjunto de datos utilizado para los experimentos que se describirán en esta sección es **Wine Data Set** [UCI], compuesto por **178 instancias**. Dichas instancias contienen **12 atributos** de carácter numérico. La **clase de destino es de carácter numérico**. Sin embargo, en este caso es de tipo entero, por lo que puede ser vista como una variable categórica. En cuanto a la metodología experimental, se ha seguido una extrategia de **HoldOut** con particionamiento de los datos de manera que $\frac{2}{3}$ son utilizados en la fase de entrenamiento y $\frac{1}{3}$ en la de test.

El experimento realizado consiste en la búsqueda de valores óptimos del kernel de las máquinas de vector soporte utilizadas. Se han utilizado kernels polinomiales (variando el grado máximo de 1 a 8) y kernels gaussianos. En el caso de los kernels lineales, los resultados se muestran en la tabla 2 de forma tabular y en la figura 2 de manera

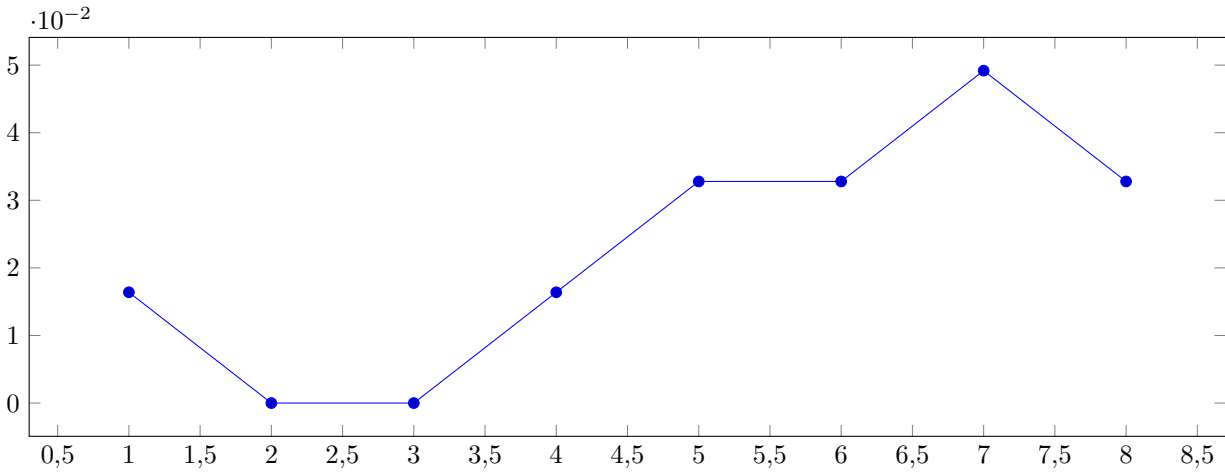


Figura 2: Tasa de error obtenida al utilizar kernels lineales en la máquina de vectores soporte sobre el conjunto de datos Wine

SVM con kernel gaussiano	
Datos	Tasa de Error
Pruebae	0,016393

Tabla 3: Tasa de error obtenida al utilizar un kernel gaussiano en la máquina de vectores soporte sobre el conjunto de datos Wine

gráfica. Tal y como se puede apreciar el óptimo se encuentra utilizando polinomios de grado 2 o 3. En este caso se preferiría el caso de polinomios de grado 2, lo cual reduce la complejidad del problema debido al número de valores a calcular.

#	Grado	Tasa de Error
1		0.016393
2		0.0
3		0.0
4		0.016393
5		0.032787
6		0.032787
7		0.04918
8		0.032787

Tabla 2: Tasa de error obtenida al utilizar kernels lineales en la máquina de vectores soporte sobre el conjunto de datos Wine

En el caso de las funciones gaussianas, la tasa de error encontrada se muestra en la tabla 3. En este caso tan solo se ha incluido un valor debido a que a pesar de probar con distintos valores para el exponente, el resultado obtenido no varía, por tanto se ha preferido no incluir los resultados de manera individual.

REFERENCIAS

- [CCAG17] Teodoro Calonge Cano and Carlos Javier Alonso González. Técnicas de Aprendizaje Automático, 2016/17.
- [GP17] Sergio García Prado. Máquinas de vectores soporte. <https://github.com/garciparedes/machine-learning-support-vector-machine>, 2017.

[too] Weka. <http://www.cs.waikato.ac.nz/ml/weka/>.

[UCI] UCI Machine Learning Repository. Wine Data Set. <http://archive.ics.uci.edu/ml/datasets/Wine>.