

MEMORIA PRÁCTICA 2

Máquinas de vector soporte

Ejercicio obligatorio:

Sergi Albiach Caro y Stéphane Díaz-Alejo León (4CO11)

Ejercicio 3.2 (obligatorio: 0.4 puntos). En el subdirectorio data/mini se encuentran dos pequeños conjuntos de datos de entrenamiento en dos dimensiones: (trSep.dat, trSeplabels.dat) y (tr.dat, trlabels.dat). El primero es linealmente separable (no es necesario kernel) y el segundo no. Para cada uno de estos conjuntos:

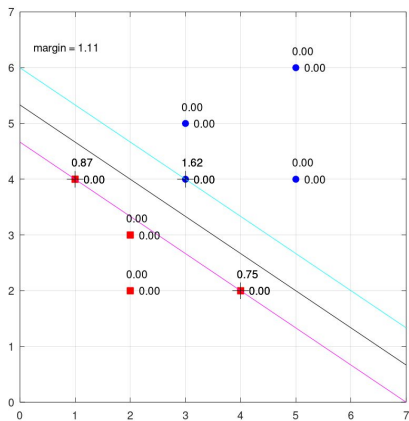
- Obtén los SVM sin kernel (es decir, kernel tipo lineal). Para simular la optimización estándar del caso separable basta usar un valor grande de C ($C = 1000$).
- Determina:
 - Los multiplicadores de Lagrange, asociados a cada dato de entrenamiento.
 - Los vectores soporte.
 - El vector de pesos y umbral de la función discriminante lineal.
 - El margen correspondiente.
- Calcula los parámetros de la frontera lineal (recta) de separación.
- Representa gráficamente los vectores de entrenamiento, marcando los que son vectores soporte, y la recta separadora correspondiente.

Además, para el conjunto no-separable utilizando diversos valores relevantes de C:

- Determina los valores de tolerancia de margen, asociados a cada dato de entrenamiento.
- Marca los vectores soporte “erróneos” en la representación gráfica.

Valores a determinar en el conjunto separable:

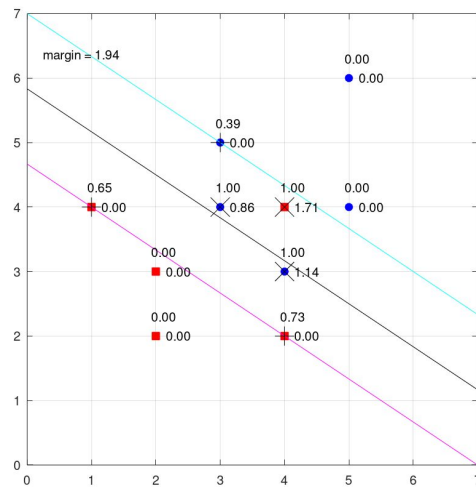
- Multiplicadores de Lagrange: [0.87472, 0.74989, -1.62461]
- Vectores soporte:
 - (1, 1) -> 1
 - (2, 1) -> 4
 - (3, 1) -> 3
 - (1, 2) -> 4
 - (2, 2) -> 2
 - (3, 2) -> 4
- Vector de pesos: [-0.99955 -1.49978]
- Umbral de la función discriminante lineal: 7.9987
- Margen: 1.1095 - Frontera lineal: $y = -(0.66647)x - (-5.3332)$



Valores a determinar en el conjunto NO separable:

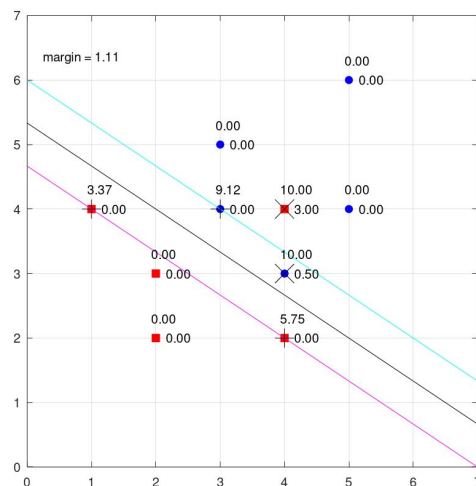
C = 1

- Multiplicadores de Lagrange: [0.65306, 0.73472, 1.0, -1.0, -0.38778, -1.0]
- Vectores soporte:
 - (1, 1) -> 1
 - (2, 1) -> 4
 - (3, 1) -> 4
 - (4, 1) -> 3
 - (5, 1) -> 3
 - (6, 1) -> 4
 - (1, 2) -> 4
 - (2, 2) -> 2
 - (3, 2) -> 4
 - (4, 2) -> 4
 - (5, 2) -> 5
 - (6, 2) -> 3
- Vector de pesos: [-0.57139, -0.85722]
- Umbral de la función discriminante lineal: 5.0003
- Margen: 1.9414
- Frontera lineal: $y = - (0.66657)*x - (-5.8331)$



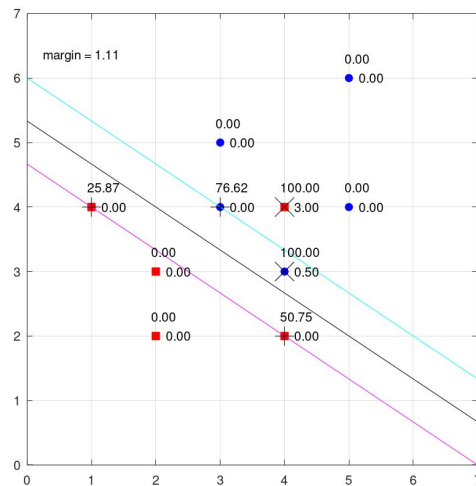
C = 10

- Multiplicadores de Lagrange: [3.3749, 5.7498, 10.0, -9.1247, -10.0]
- Vectores soporte:
 - (1, 1) -> 1
 - (2, 1) -> 4
 - (3, 1) -> 4
 - (4, 1) -> 3
 - (5, 1) -> 4
 - (1, 2) -> 4
 - (2, 2) -> 2
 - (3, 2) -> 4
 - (4, 2) -> 4
 - (5, 2) -> 3
- Vector de pesos: [-1.0 -1.4996]
- Umbral de la función discriminante lineal: 7.9986
- Margen: 1.1096
- Frontera lineal: $y = -(0.66682)*x - (-5.3336)$



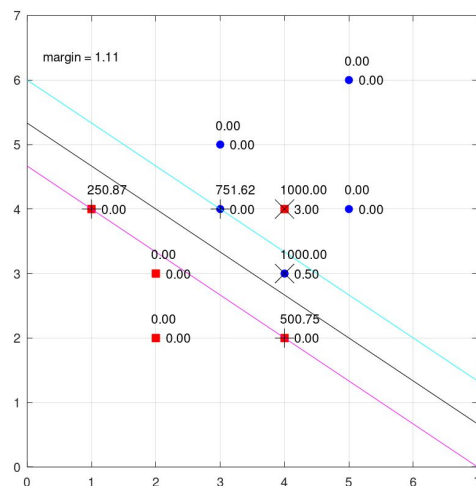
C = 100

- Multiplicadores de Lagrange: [25.875, 50.750, 100.0, -76.625, -100.0]
- Vectores soporte:
 - (1, 1) -> 1
 - (2, 1) -> 4
 - (3, 1) -> 4
 - (4, 1) -> 3
 - (5, 1) -> 4
 - (1, 2) -> 4
 - (2, 2) -> 2
 - (3, 2) -> 4
 - (4, 2) -> 4
 - (5, 2) -> 3
- Vector de pesos: [-0.99955 -1.49978]
- Umbral de la función discriminante lineal: 7.9987
- Margen: 1.1095
- Frontera lineal: $y = -(0.66647)*x - (-5.3332)$



C = 1000

- Multiplicadores de Lagrange: [250.87, 500.75, 1000.0, -751.62, -1000.0]
- Vectores soporte:
 - (1, 1) -> 1
 - (2, 1) -> 4
 - (3, 1) -> 4
 - (4, 1) -> 3
 - (5, 1) -> 4
 - (1, 2) -> 4
 - (2, 2) -> 2
 - (3, 2) -> 4
 - (4, 2) -> 4
 - (5, 2) -> 3
- Vector de pesos: [-0.99955 -1.49977]
- Umbral de la función discriminante lineal: 7.9986
- Margen: 1.1097
- Frontera lineal: $y = -(0.66647)*x - (-5.3332)$



Ejercicio 3.3 (obligatorio: 0.4 puntos). Realiza un experimento donde se evalúe el error de clasificación en función de los parámetros del clasificador basado en SVM. Más concretamente, explora los valores del parámetro C (-c 1, 10, 100...) y el tipo de de kernel (-t 0, 1, 2, 3). Para aquellos tipos de kernel que lo permitan, explora sus parámetros específicos. Por ejemplo, en el caso del kernel polinomial (-t 1), explora el grado del polinomio (-d 1, 2, 3, 4, 5).

En función del número de resultados obtenidos como consecuencia de la exploración de los valores de los parámetros, utiliza una representación adecuada de los mismos, ya sea gráfica o tabular, que muestre no solo el mejor resultado obtenido, sino también otros resultados relevantes que permitan poner de manifiesto la tendencia a mejorar o empeorar del modelo según varían los valores de los parámetros considerados.

Tipo de Kernel	C	Grado	Precisión
Lineal	1	-	84.7167%
Lineal	10	-	84.7167%
Lineal	100	-	84.7167%
Lineal	1000	-	84.7167%
Lineal	10000	-	84.7167%
Polinomial	1	1	92.0667%
Polinomial	1	2	98.0500%
Polinomial	1	3	97.8833%
Polinomial	1	4	97.3500%
Polinomial	1	5	96.2167%
Polinomial	10	1	91.8833%
Polinomial	10	2	98.0500%
Polinomial	10	3	97.8833%
Polinomial	10	4	97.3500%
Polinomial	10	5	96.2167%
Polinomial	100	1	86.2500%
Polinomial	100	2	98.0500%

Polinomial	100	3	97.8833%
Polinomial	100	4	97.3500%
Polinomial	100	5	96.2167%
Polinomial	1000	1	85.5500%
Polinomial	1000	2	98.0500%
Polinomial	1000	3	97.8833%
Polinomial	1000	4	97.3500%
Polinomial	1000	5	96.2167%
Polinomial	10000	1	85.5500%
Polinomial	10000	2	98.0500%
Polinomial	10000	3	97.8833%
Polinomial	10000	4	97.3500%
Polinomial	10000	5	96.2167%
Radial	1	-	11.7167%
Radial	10	-	11.6333%
Radial	100	-	11.6333%
Radial	1000	-	11.6333%
Radial	10000	-	11.6333%
Sigmoide	1	-	11.6333%
Sigmoide	10	-	11.6167%
Sigmoide	100	-	11.6167%
Sigmoide	1000	-	11.6167%
Sigmoide	10000	-	11.6167%

Los mejores resultados se obtienen para el kernel polinomial con C=1 - grado=2, C=10 - grado=2, C=100 - grado=2 (98,05%). Para el kernel lineal vemos que la precisión es constante (84.7167%) y que, para el sigmoide y radial, la precisión cae al 11.6% (excepto Radial C=1 que es 11.7%) independientemente de la C.

Ejercicio 3.4 (obligatorio: 0.2 puntos). Una vez determinado los valores óptimos de los parámetros del clasificador basado en SVM, entrena y evalúa un clasificador final en los conjuntos oficiales MNIST de entrenamiento y test, respectivamente. Para ello te recomendamos que tomes como punto de partida el script `svm-exp.m`, modificándolo adecuadamente para generar el script `svm-eva.m` que también deberá recibir como entrada el conjunto de test de MNIST. Recuerda que toda estimación de (la probabilidad de) error de un clasificador final, debe ir acompañada de sus correspondientes intervalos de confianza al 95%. Discute los resultados obtenidos comparándolos con los obtenidos con el clasificador de mixtura de gaussianas y con los reportados en la tarea MNIST, especialmente los basados en SVM.

Con los valores óptimos ($t=1$, $d=2$), hemos obtenido los siguientes resultados variando el parámetro C para observar si existían diferencias notables:

Tipo de Kernel	C	Grado	Precisión	Inter. Confianza
Polinomial	1	2	98.050%	[97.779 98.321]
Polinomial	10	2	98.050%	[97.779 98.321]
Polinomial	100	2	98.050%	[97.779 98.321]
Polinomial	1000	2	98.050%	[97.779 98.321]
Polinomial	10000	2	98.050%	[97.779 98.321]

Cómo se puede observar, con cualquiera de los valores del parámetro C se obtiene la misma precisión, 98.05%, lo que nos deja con un porcentaje de error del 1.95%. Este porcentaje se acerca al proporcionado en la página web de MNIST, en la cual, con un SVM de Kernel Gaussiano se obtiene un porcentaje de error del 1.4%.