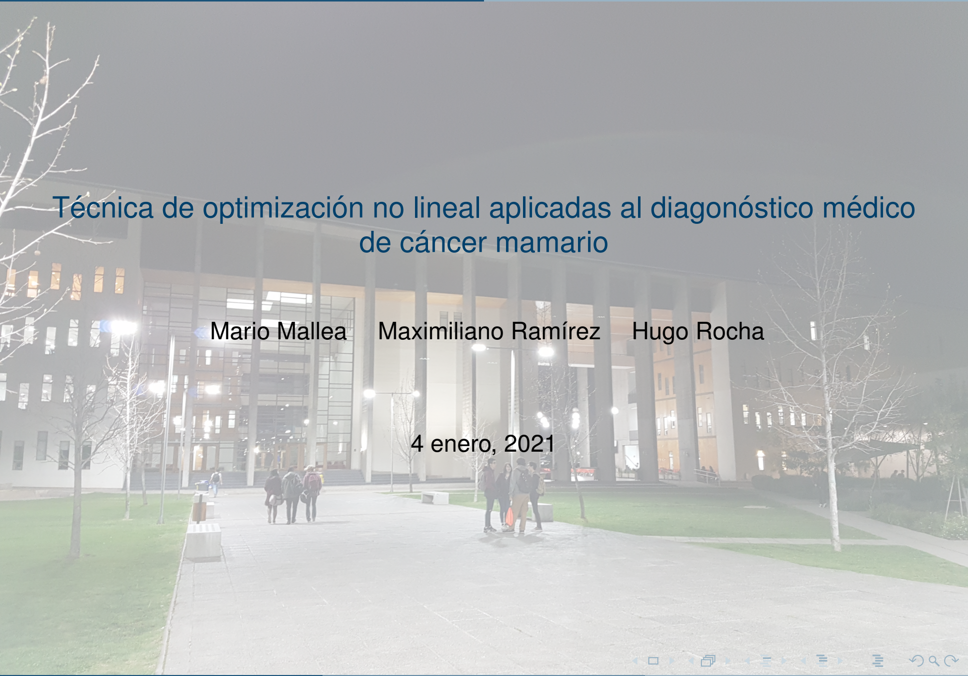


Técnica de optimización no lineal aplicadas al diagonóstico médico de cáncer mamario

Mario Mallea Maximiliano Ramírez Hugo Rocha

4 enero, 2021

Indice

Introduction

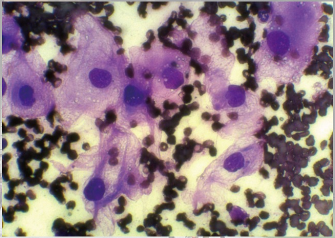
- A continuación se presenta nuestro trabajo sobre técnicas de optimización no lineal aplicados a la detección del cáncer mamario a partir de imágenes de biopsias de células tumorales. Logrando obtener una caracterización de la solución al problema de clasificación y por tanto una posible automatización del diagnóstico.

Motivación

El Cáncer en Cifras

A nivel mundial en el año 2012 cerca de 521.000 mujeres murieron a causa del cáncer de mama[a].

[a] Global burden of cancer in womans



Biopsia con aguja fina de masa mamaria con presencia de células tumorales

Formulación como problema de optimización.

Hiperplano separador

- Un hiperplano separador es una superficie de codimensión 1[a] de decisión que separa las distintas clases que tiene el problema de clasificación. El hiperplano sera denotado por $\mathcal{H}$ es de la forma:

$$w^T x + b = 0$$

[a] es decir, es un subespacio de una dimensión menos que el espacio en el que se trabaja

Margen del Hiperplano

- El margen geométrico de un hiperplano $\mathcal{H}$ con respecto a la data, se define como la distancia más corta de un punto de entrenamiento x_i al hiperplano $\mathcal{H}$.

Hiperplano óptimo

- Consideraremos como hiperplano separador óptimo $\mathcal{H}$ aquel que tiene el mayor margen posible

Formulación como problema de optimización

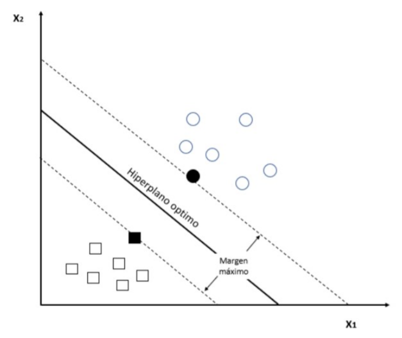


Figure 1: Construcción del margen.

Formulación como problema de optimización

Formulación Formal del problema

$$
\begin{cases}
\displaystyle \max_{w} \min_{x_i} \left| \langle x_i, \frac{w}{||w||} \rangle + b \right| \\[2ex]
\textbf{s.t:} \quad sign(\langle x_i, w \rangle + b) = sign(y_i)
\end{cases}
$$

Formulación como problema de optimización

SVM duro

$$\begin{cases} \quad min \quad \dfrac{1}{2}||w||^2 \\ \textbf{s.t:} \quad (\langle x_i, w\rangle + b)y_i \geq 1 \end{cases} \tag{1}$$

Formulación como problema de optimización.

Problema dual duro

- Veamos ahora que es posible obtener la formulación dual de nuestro problema de optimización dado por:

$$
\begin{aligned}
\max_{\alpha_i \geq 0} \quad & \sum_{i}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\
\text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i = 1, 2, ..., n, \\
& \sum_{i=1}^{n} \alpha_i y_i = 0 \quad \forall i = 1, ..., n
\end{aligned} \tag{2}
$$

Formulación como problema de optimización.

¿Que hacer si se nos presente el siguiente caso?

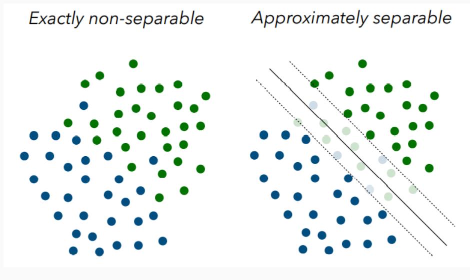


Figure 2: Caso casi linealmente separable.

- Consideraremos que nuestros datos son casi linealmente separables si salvo una cantidad reducida con respecto a la data nuestro problema es linealmente separable.

Formulación como problema de optimización.

SVM blando

$$
\begin{aligned}
\min_{w,b,\xi} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i \\
\text{s.t.} \quad & (\langle x_i, w\rangle + b)y_i \geq 1 - \xi_i, \\
& \qquad\qquad\qquad \xi_i \geq 0 \forall i = 1, ..., n
\end{aligned}
\tag{3}
$$

Formulación como problema de optimización.

Formulación SVM dual blando

$$\begin{aligned} \max_{\alpha_i \geq 0} \quad & \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s.t.} \quad & C \geq \alpha_i \geq 0 \quad \forall i = 1, 2, ..., n, \\ & \sum_{i=1}^{n} \alpha_i y_i = 0 \quad \forall i = 1, ..., n \end{aligned} \tag{4}$$

Observación

- $$\sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle = \frac{1}{2} \alpha^T \Lambda \alpha + q^T \alpha \tag{5}$$

  se tendrá que la matriz $\Lambda \in \mathbb{R}^N \times \mathbb{R}^N$ es definida positiva, así la función a maximizar es cóncava, por lo que se tiene que el máximo del problema es único.

Formulación como problema de optimización.

¿Que hacemos si ahora se nos presenta el siguiente caso?

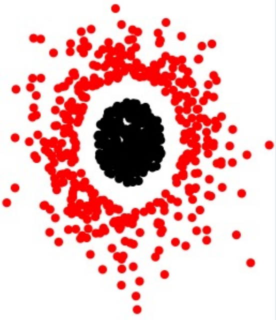


Figure 3: Caso no linealmente separable.

- Consideraremos que nuestros datos son no linealmente separable cuando bajo ningún punto de vista es posible separar nuestros datos por medio de un hiperplano.

Formulación como problema de optimización.

Teorema de Cover

- Sea $\phi : \mathbb{R}^d \to \mathbb{R}^D$. Para cualquier conjunto de $n \in \mathbb{Z}$, ejemplos $S = \{(x^{(l)}, y^{(l)})\}_{l=1}^{n}$, $x^{(l)} \in \mathbb{R}^d$, $y^{(l)} \in \{\pm 1\}$, tenemos que $\phi(S) = \{(\phi(x^{(l)}), y^{(l)})\}$ es linealmente separable con probabilidad:

$$P(n, D) = \frac{1}{2^{n-1}} \sum_{k=0}^{D-1} \binom{n-1}{k}$$

Formulación como problema de optimización.

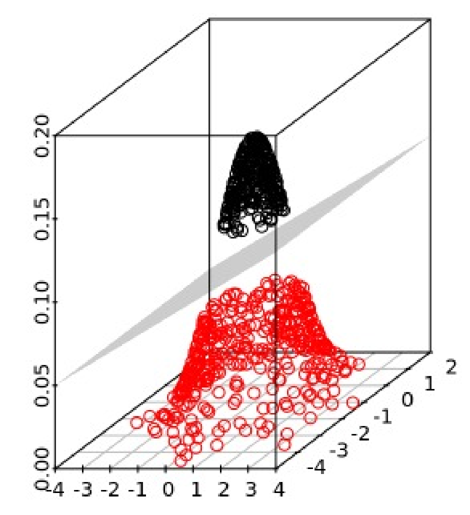


Figure 4: Caso linealmente separable en espacio de mayor dimensionalidad.

Formulación como problema de optimización.

Función Kernel

- Denominaremos funciones Kernel, a aquellas que cumplan con lo siguiente:

$$\kappa(x, y) = \langle \phi(x), \phi(y) \rangle \tag{6}$$

Dónde ϕ es una función que va de X a un espacio de mayor dimensionalidad dotado de un producto interno.

Formulación como problema de optimización.

SVM dual blando Kernelizado

$$
\begin{aligned}
\max_{\alpha_i \geq 0} \quad & \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\
\text{s.t.} \quad & C \geq \alpha_i \geq 0 \quad \forall i = 1, 2, ..., n, \\
& \sum_{i=1}^{n} \alpha_i y_i = 0 \quad \forall i = 1, ..., n
\end{aligned} \tag{7}
$$

Formulación como problema de optimización

Algunos Kernels populares

- Lineal:

$$k(x, x') = x^T x'$$

- RBF o Gaussiano de escala γ:

$$k(x, x') = e^{-\gamma ||x^T x'||^2}$$

- Polinomial de grado p, intercepto c_0 y escala γ:

$$k(x, x') = (\gamma x^T x' + c_0)^p$$

- Laplaciano de escala γ:

$$k(x, x') = e^{-\gamma ||x^T x||_1}$$

Análisis del problema de optimización

Análisis de sensibilidad

- Problema de programación matemática perturbado

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq u_i \qquad i = 1, ..., m \end{aligned} \tag{8}$$

Análisis del problema de optimización

Análisis de sensibilidad

- Condición de optimalidad del problema perturbado

$$p^*(u) \geq p^*(0) - {\alpha^*}^T u \tag{9}$$

Análisis de sensibilidad

- α_i^* grande y $u_i < 0 \Longrightarrow p^*(u)$ aumentará demasiado
- α_i^* pequeño y $u_i > 0 \Longrightarrow p^*(u)$ no disminuye en gran medida

Observación

A partir de (9) es posible deducir que la sensibilidad del problema viene dada por la magnitud del multiplicador de Lagrange asociado a cada restricción

Análisis de sensibilidad

SVM dual blando Kernelizado

$$
\begin{aligned}
\max_{\alpha_i \geq 0} \quad & \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\
\text{s.t.} \quad & C \geq \alpha_i \geq 0 \quad \forall i = 1, 2, ..., n, \\
& \sum_{i=1}^{n} \alpha_i y_i = 0 \quad \forall i = 1, ..., n
\end{aligned} \tag{10}
$$

Aproximación de support vectors

Función de decisión

$$F(x) = sign\left(\sum_i(\alpha_i y_i \kappa(x_i, x)) + b^*\right) \tag{11}$$

Dónde

$$\begin{aligned} b^* &= |B|^{-1} \sum_{\ell} \left(y^{(\ell)} - w^T x^{(\ell)}\right) \\ &= |B|^{-1} \sum_{\ell} \left(y^{(\ell)} - \sum_{\ell'} \alpha_{\ell'} y^{(\ell')} \left\langle x^{(\ell')} x^{(\ell)} \right\rangle\right) \end{aligned} \tag{12}$$

Observación

La función de decisión sólo depende de los multiplicadores de Lagrange no nulos.

Rutinas de Resolución

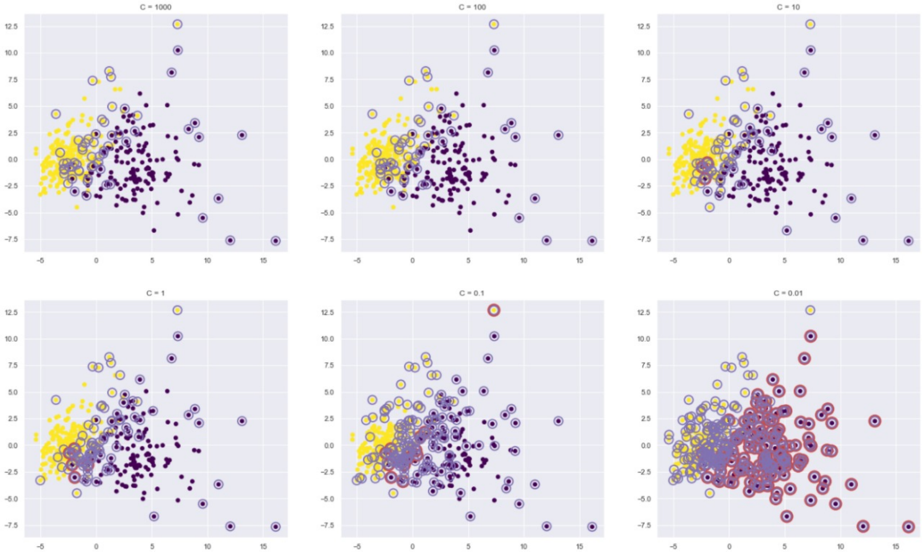
Problema de programación cuadrática

$$\begin{aligned} \min \quad & \frac{1}{2}\alpha^T P\alpha + q^T\alpha \\ \text{s.t.} \quad & G\alpha \leq h, \\ & A\alpha = b \end{aligned} \tag{13}$$

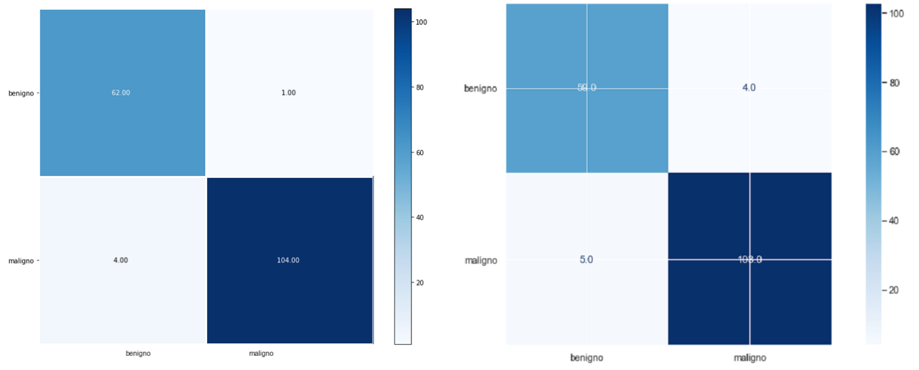
Problema dual blando como uno de programación cuadrática

$$\begin{aligned} \max_{\alpha_i \geq 0} \quad & \sum_{i}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \kappa(x_i, x_j) \\ \text{s.t.} \quad & C \geq \alpha_i \geq 0 \quad \forall i = 1, ..., n, \\ & \sum_{i=1}^{n} \alpha_i y_i = 0 \quad \forall i = 1, ..., n \end{aligned} \tag{14}$$

Representación y caracterización de soluciones



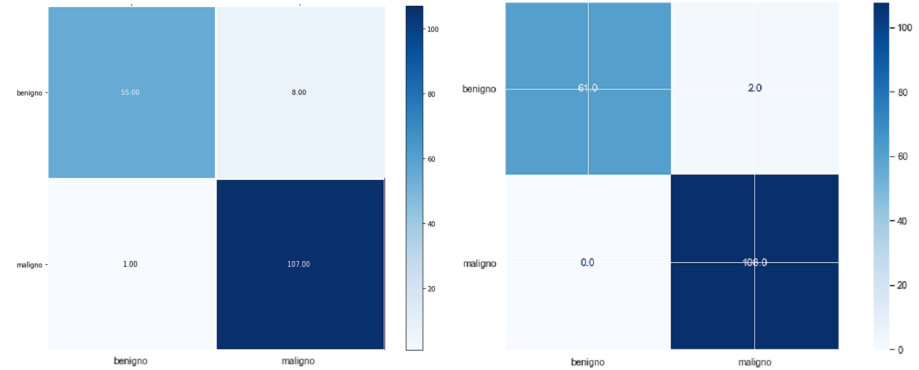
Representación y caracterización de soluciones



(a) QP (b) SVM

Figure 5: Matriz de Confusión: Kernel Lineal y regularización =100

Representación y caracterización de soluciones



(a) QP (b) SVM

Figure 6: Matriz de Confusión: Kernel Gaussiano y regularización =100

Conclusión

- Automatización del diagnóstico $\Longrightarrow$ grandes cantidades de datos.
- Ineficiencia en memoria para QP.
- SVM tiene lenta velocidad de convergencia.
- Solución: Reducción de datos $\Longrightarrow$ Propuesta: Estimar.

Gracias