INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE OCCIDENTE

MCD3395A – OPTIMIZACIÓN CONVEXA

PRIMAVERA 2022

ITESO, Universidad Jesuita de Guadalajara

MÓDULO 2 – ACTIVIDAD 1

HW 05: SUPPORT VECTOR MACHINE

Equipo:

Leonardo Razo Islas

Gabriela Lozano Orozco

Emilio Carranza Ávila

Daniel Francisco Nuño Álvarez

Tlaquepaque, Jalisco. Febrero de 2022.

Problem 1: LS-SVM: Regression

Consider the following optimization problem:

$$\begin{aligned} \min_{w,b,e} \mathcal{P}(w,e) &= \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2 \\ \text{s. t. } y_k &= w^T\varphi(x_k) + b + e_k, \quad k = 1,\ldots,N. \end{aligned} \tag{1}$$

where $\{x_k, y_k\}_{k=1}^N$ represents a training set with input data $x_k \in \mathbb{R}^n$, the output data given $y_k \in \mathbb{R}$, $e_k \in \mathbb{R}^n$ are slack variables, and the feature maps have the form $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$. Then, the model's parameters are $w \in \mathbb{R}^m$ and $b \in \mathbb{R}$. Finally, $\gamma > 0$.

Note that the problem (1) can be written as:

$$\min_{w,b,e} \mathcal{P}(w,e) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{k=1}^{N}\left[y_k - \left(w^T\varphi(x_k) + b\right)\right]^2.$$

Thus, the problem (1) is related to the so-called least squares support vector machines LS-SVM.

1. Show that the Lagrangian of the problem (1) is given by:

$$\mathcal{L}(w,b,e;\alpha) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N}\alpha_k\left\{w^T\varphi(x_k) + b + e_k - y_k\right\}$$

2. Since the problem has not inequality constraints, the KKT optimality follows directly from the first-order conditions provided by the gradient of the Lagrangian $\mathcal{L}(w,b,e;\alpha)$. Then, show that:

   - $\nabla_w\mathcal{L} = 0$ implies $w = \sum_{k=1}^N \alpha_k\varphi(x_k)$.
   - $\frac{\partial\mathcal{L}}{\partial b} = 0$ implies $\sum_{k=1}^N \alpha_k = 0$.
   - $\frac{\partial\mathcal{L}}{\partial e_k} = 0$ implies $\alpha_k = \gamma e_k$ for $k = 1,\ldots,N$.
   - $\frac{\partial\mathcal{L}}{\partial \alpha_k} = 0$ implies $w^T\varphi(x_k) + b + e_k - y_k = 0$ for $k = 1,\ldots,N$.

3. Define adequate vector variables, such that the optimization problem reduces to a set of linear equations which must be solved for α and b.

Tarea 5

Prob 1
Regresión

$$\min_{w,b,e} P(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$$

$$\text{s.t. } y_k = w^T \varphi(x_k) + b + e_k \quad , \quad k = 1, \ldots, N$$

donde: $\{x_k, y_k\}_{k=1}^{N}$ es un Dataset de entrenamiento

con datos $x_k \in \mathbb{R}^n$ y el resultado $y_k \in \mathbb{R}$

$e_k \in \mathbb{R}^n$ } error

$\varphi(\cdot): \mathbb{R}^n \to \mathbb{R}^m$ } espacio de características

$w \in \mathbb{R}^m$, $b \in \mathbb{R}$ } parámetros del modelo

$\gamma > 0$ } Factor de regularización

1) muestra el lagrangeano de P → igualado a cero.

es reescribir P, agregando la restricción multiplicada por el multiplicador de lagrange, en este caso: α_k

I

$$\boxed{L(w,b,e;\alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k \underbrace{(w^T \varphi(x_k) + b + e_k - y_k)}}$$

resulta de igualar a cero la restricción

2) condiciones de 1er orden de $L(w,b,e;\alpha)$

II

$$\nabla_w L = w + 0 - \sum_{k=1}^{N} \alpha_k \varphi(x_k) = 0 \quad y \therefore \quad \boxed{w = \sum_{k=1}^{N} \alpha_k \varphi(x_k)}$$

$\frac{1}{2} w^T w = \frac{1}{2}(w_1^2 + w_2^2 + \ldots w_N^2)$, la derivada de w_k

$\frac{\cancel{2} w_k}{\cancel{2}} = w_k$ y así para cada k, entonces se forma vector $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}$

$$\frac{\partial L}{\partial b} = 0 + 0 - \sum_{k=1}^{N} \alpha_k = 0 \quad \therefore \quad \boxed{\sum_{k=1}^{N} \alpha_k = 0}$$ III

$$\frac{\partial L}{\partial e_k} = 0 + \gamma e_k - \alpha_k \quad , k = 1, \dots N \quad \therefore \quad \boxed{e_k = \frac{\alpha_k}{\gamma}}$$ IV

$$\frac{\partial L}{\partial \alpha_k} = 0 + 0 + w^T \varphi(x_k) + b + e_k - y_k = 0 \quad , \quad k = 1, \dots N$$

$$\boxed{w^T \varphi(x_k) + b + e_k - y_k = 0}$$ V

3) Reducir sistema de ecns II III IV y V a un sistema lineal para α y b

sustituir II y IV en V

$$\boxed{y_k = \sum_{l=1}^{N} \alpha_l \varphi^T(x_l) \varphi(x_k) + b + \frac{\alpha_k}{\gamma}}$$ VI

abriendo VI y encontrando patrón para reescribir con vectores

$$K = \begin{bmatrix} \varphi^T(x_1)\varphi(x_1) & \varphi^T(x_2)\varphi(x_1) & \dots & \varphi^T(x_N)(x_1) \\ \varphi^T(x_1)\varphi(x_2) & \varphi^T(x_2)\varphi(x_2) & \dots & \varphi^T(x_N)(x_2) \\ \vdots & & & \\ & & & \varphi^T(x_N)(x_N) \\ \varphi^T(x_1)\varphi(x_N) & \dots & \dots & \end{bmatrix}$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \quad 1_v = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$K\alpha + b 1_v + \frac{1}{\gamma}\alpha = Y$$

sacando fact. común α

$$\boxed{\left(K + \frac{1}{\gamma} I\right)\alpha + b 1_v = Y}$$

considerando III $\sum \alpha_k = 0$

reescribiendo $\boxed{1_v^T \alpha = 0}$

sistema lineal

$$\begin{bmatrix} 1_v^T & 0 \\ K + 1/\gamma & 1_v \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}$$

despejando para parámetros α y b

$$\boxed{\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 1_v^T & 0 \\ K + 1/\gamma & 1_v \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ Y \end{bmatrix}}$$

4. Finally, explain (with math) how the optimization problem (1) is related to the regression problem.

1.4 Siendo el kernel $K(x_k, x)$,

haciendo la múltiplicación de matrices

$$\sum 0 \cdot b + \sum \vec{1}_v^T \alpha = 0$$

$$\sum \vec{1}_v \cdot b + \sum (K(x_k, x) + \gamma^{-1} I) \cdot \alpha_k = Y$$

la función estimación del modelo queda como una regesión

$$y(x) = \sum \alpha_k K(x_k, x) + b$$

Problem 2: LS-SVM: Classification

Consider the following optimization problem:

$$\begin{aligned} &\min_{w,b,e} \mathcal{P}(w,e) = \frac{1}{2}w^T w + \gamma \frac{1}{2}\sum_{k=1}^{N} e_k^2 \\ &\text{s. t.} \quad y_k \left[w^T \varphi(x_k) + b \right] = 1 - e_k, \quad k = 1, \ldots, N. \end{aligned} \tag{2}$$

where $y_k \in \{-1, 1\}$ is the response (target) variable.
Then conduct an analysis of the problem (2) by applying the steps (a) to (d) of the problem (1) with, possibly, their respective modifications. In the step (d), explain how the new problem (2) is related to the classification problem. Finally, compare KKT matrix system obtained for this case with that of the problem (1).

Prob 2

clasificación

$$\min_{w,b,e} P(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$$

$$\text{s.t.} \quad y_k (w^T \varphi(x_k) + b) = 1 - e_k \qquad k = 1, \dots N$$

Donde $y_k \in \{-1, 1\}$ es la variable respuesta

Similar a prob 1, pero ahora la salida es categórica $< \begin{matrix} 1 \\ -1 \end{matrix}$

1) lagrangeano

restricción $\rightarrow 0 = y_k (w^T \varphi(x_k) + b) - 1 + e_k$

$$\boxed{L(w,b,e;\alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 - \sum_{k=1}^{N} \alpha_k \left[y_k (w^T \varphi(x_k) + b) - 1 + e_k \right]}$$

2) condiciones de 1er orden de $L(w,b,e;\alpha)$

$$\nabla_w L = w + 0 - \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k) = 0 \therefore \boxed{w = \sum_{k=1}^{N} \alpha_k y_k \varphi(x_k)} \quad \text{I}$$

$$\frac{\partial L}{\partial b} = 0 + 0 - \sum_{k=1}^{N} \alpha_k y_k = 0 \quad \therefore \boxed{\sum_{k=1}^{N} \alpha_k y_k = 0} \quad \text{II}$$

$$\frac{\partial L}{\partial e_k} = 0 + \gamma e_k - \alpha_k, \quad k = 1 \dots, N \qquad \boxed{e_k = \frac{\alpha_k}{\gamma}} \quad \text{III}$$

$$\frac{\partial L}{\partial \alpha_k} = 0 + 0 + y_k (w^T \varphi(x_k) + b) - 1 + e_k = 0, \quad k = 1, \dots, N$$

$$\boxed{y_k (w^T \varphi(x_k) + b) - 1 + e_k = 0} \quad \text{IV}$$

3) Reducir sist. ecns. ?

sustituir I y III en IV

$$\boxed{y_k \left[\sum_{l=1}^{N} \alpha_l y_k \varphi^T(x_l) \varphi(x_k) + b \right] - 1 + \frac{\alpha_k}{\gamma} = 0} \quad \text{V}$$

abriendo V

$$K = \begin{bmatrix} \cancel{y_1} \varphi^T(x_1)\varphi(x_1) & \cancel{y_1}\varphi^T(x_2)\varphi(x_1) & \cdots & y_1\varphi^T(x_N)\varphi(x_1) \\ \cancel{y_2}\varphi^T(x_1)\varphi(x_2) & \cancel{y_2}\varphi^T(x_2)\varphi(x_2) & \cdots & y_2\varphi^T(x_N)\varphi(x_2) \\ \vdots & & & \vdots \\ y_N\varphi^T(x_1)\varphi(x_N) & y_N\varphi^T(x_2)\varphi(x_N) & \cdots & y_N\varphi^T(x_N)\varphi(x_N) \end{bmatrix}$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \qquad 1_v = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$Y\left[\alpha Y K + b 1_v\right] - 1 + \frac{\alpha}{\gamma} = 0$$

$$\alpha Y Y^T K + Y b 1_v - 1 + \frac{\alpha}{\gamma} = 0$$ sacar factor común α y acomodando en términos de α y b

$$\boxed{\alpha\left(YY^TK + \frac{1}{\gamma}I\right) + Yb = +1}$$

considerando II $\quad \sum \alpha_K y_K = 0 \implies$ reescribir $\boxed{\alpha y^T = 0}$

$$y(b) + \alpha(Z t^T + \gamma^{-1} I) = 1_v$$

$\square$ y $\square$ forman sistema lineal

$$\begin{bmatrix} Y^T & 0 \\ YY^TK + \frac{1}{\gamma}I & Y \end{bmatrix}\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

despejando para α y b

$$\boxed{\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} Y^T & 0 \\ YY^TK + \frac{1}{\gamma}I & Y \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix}}$$

Problem 3: A LS-SVM Formulation of the PCA

Consider the following optimization problem:

$$\begin{aligned} &\min_{w,b,e} \mathcal{P}(w,e) = \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 - \frac{1}{2} w^T w \\ &\text{s. t. } e_k = w^T x_k, \quad k = 1, \dots, N. \end{aligned} \tag{3}$$

where $\{x_k\}_{k=1}^N$ with input data $x_k \in \mathbb{R}^n$ represents an unlabeled data-set.

1. Calculate the Langragian for the problem (3).
2. Show that the KKT matrix system (the dual problem) is given by the eigenvalue problem:

$$\begin{bmatrix} x_1^T x_1 & \dots & x_1^T x_N \\ \vdots & & \vdots \\ x_N^T x_1 & \dots & x_N^T x_N \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = \lambda \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$$

   with $\lambda = 1/\gamma$.

3. Finally, explain (with math) how the optimization problem (3) is related to the principal component analysis (PCA) method.

Prob 3

PCA

$$\min_{w,b,e} P(w,e) = \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 - \frac{1}{2} w^T w$$

?

s.t. $e_k = w^T x_k$, $k = 1, \dots, N$

Donde $\{x_k\}_{k=1}^{N}$ entradas donde $x_k \in \mathbb{R}^n$ son datos sin etiquetar

1) lagrangeano

$$\boxed{L(w,e;\alpha) = \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 - \frac{1}{2} w^T w - \sum_{k=1}^{N} \alpha_k (w^T x_k - e_k)}$$

2) KKT matrices

$$\nabla_w L = 0 - w - \sum_{k=1}^{N} \alpha_k x_k = 0 \quad \therefore \quad \boxed{w = -\sum_{k=1}^{N} \alpha_k x_k} \quad \text{I}$$

$$\frac{\partial L}{\partial e_k} = \gamma e_k - 0 + \alpha_k = 0 \quad \therefore \quad \boxed{e_k = \frac{-\alpha_k}{\gamma}} \quad \text{II} \quad k = 1, \dots N$$

$$\frac{\partial L}{\partial \alpha_k} = 0 - 0 - w^T x_k + e_k = 0 \quad k = 1, \dots, N \quad \therefore$$

$$\boxed{e_k - w^T x_k = 0} \quad \text{III}$$

3ecn y 3incógnitas (w, α_k, e_k)

I y II en III

$$\boxed{\frac{-\alpha_k}{\gamma} + \sum_{l=1}^{N} \alpha_l x_l \, x_k = 0} \quad \text{IV}$$

abriendo y encontrando patrones para escribir como vectores

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad K = \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ x_1 & x_2 & \dots & x_N \\ \vdots \\ x_1 & x_2 & \dots & x_N \end{bmatrix}$$

$$-\sum \alpha_\ell x_\ell x_k = \alpha_k/\Gamma$$

$$K_1 \quad \sum \alpha_\ell x_\ell x_1 = \alpha_1/\Gamma$$
$$K_2 \quad \sum \alpha_\ell x_\ell x_2 = \alpha_2/\Gamma$$
$$K_n \quad \sum \alpha_\ell x_\ell x_n = \alpha_n/\Gamma$$

$$\begin{array}{cccc}
\ell_1 & \ell_2 & \ell_n & \\
\alpha_1 x_1 x_1 & \alpha_2 x_2 x_1 & \alpha_n x_n x_1 & = \alpha_1/\Gamma \\
\alpha_1 x_1 x_2 & \alpha_2 x_2 x_2 & \alpha_n x_n x_2 & = \alpha_2/\Gamma \\
\alpha_1 x_1 x_n & \alpha_2 x_2 x_n & \alpha_n x_n x_n & = \alpha_n/\Gamma
\end{array}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_n \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_n \end{bmatrix} = 1/\Gamma \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_n \end{bmatrix}$$

2

$$-\alpha\left(\frac{1}{\gamma}\right) + \alpha K x = 0$$

$\boxed{\alpha K x = \frac{1}{\gamma}\alpha}$ Despejar para conoce α

3.3

23 Feb, 2022

La formulación de principal component Analysis establece $\{x_k\}_{k=1}^{N}$ con $x_k \in \mathbb{R}^n$ encuentra las variables $W^T X_k$ con la maxima varianza

$$\max_W Var(W^T x) = cov(W^T X, W^T X) \simeq \frac{1}{N}\sum_{k=1}^{n}(W^T X_k)^2$$
$$= W^T C W$$

donde $C = \frac{1}{N}\sum_{k=1}^{n} x_k x_k^T$ y $W^T W = 1$ ⇒ la sumatoria de los pesos al cuadrado $=1$ de otra manera los pesos podrían ser asignados arbitrariamente para maximizar la varianza.

es una optimización restringida

$$L(w;\lambda) = \frac{1}{2} w^T C w - \lambda(w^T w - 1)$$

siendo λ el multiplicador de Lagrange

Esencialmente, PCA es una combinacion lineal de los eigenvectores que maximizan la varianza usando la matriz de covarianza

Para los eigenvalores $C_w = \lambda_w$

con $C = C^T \geq 0$, obtenido de $\frac{\partial L}{\partial w} = 0$, $\frac{\partial L}{\partial \lambda} = 0$

Los principal componet scores, denotados por $z := W^T x$

se convierte en

$$Z(x) = W^T x = \sum_{l=1}^{N} \alpha_l x_l^T x$$

La solucion optima correspondiente al eigenvalor más grande

$$\sum (W^T X_k)^2 = \sum_{k=1}^{N} e_k^2 = \sum_{k=1}^{N} \frac{1}{\gamma^2} \sigma_k^2 = \lambda_{max}^2$$

donde $\sum_{k=1}^{N} \alpha_k^2 = 1$ para el eigenvector normalizado