ITESO, Universidad Jesuita de Guadalajara

Maestría de Ciencia de Datos

Optimización Convexa

Tarea 4: Optimización restringida y regularización

Estudiante: Daniel Nuño
Profesor: Dr. Juan Diego Sanchez Torres
Fecha entrega: 16 de febrero, 2022

Introduction

The purpose of this last activity is to solve a problem similar to those encountered in practice, using a reasonably sophisticated code structure appropriate to the solution of the problem. Also, use is made of symbolic transformers for the resolution of a regression problem. It is demonstrated that the use of new inputs can improve the model's predictive capacity.

Problema 1

Activities

Problem 1: Solve the following constrained optimization problems.

1. Maximize $f(x,y) = \sqrt{6-x^2-y^2}$ subject to the constraint, $x+y-2=0$.
2. Maximize $U(x,y) = 8x^{4/5}y^{1/5}$ subject to the constraint, $4x+2y=12$.
3. Optimize $f(x,y,z) = yz + xy$ subject to the constraints: $xy=1, \quad y^2+z^2=1$.
4. Minimize $f(x,y,z) = x^2+y^2+z^2$ when $x+y+z=9$ and $x+2y+3z=20$.
5. A large container in the shape of a rectangular solid must have a volume of 480 m^3. The bottom of the container costs $\$5/\text{m}^2$ to construct whereas the top and sides cost $\$3/\text{m}^2$ to construct. Use Lagrange multipliers to find the dimensions of the container of this size that has the minimum cost.
6. Use Lagrange multipliers to find the point on the line $y = 2x+3$ that is closest to point $(4,2)$.
7. Use Lagrange multipliers to find the minimum distance from point $(0,1)$ to the parabola $x^2=4y$.
8. Use Lagrange multipliers to find the minimum and maximum distances between the ellipse $x^2+xy+2y^2 = 1$ and the origin.

Lagrange dice que hay un número (lambda) que hace los vectores gradientes paralelos (de la función a optimizar y la restricción).

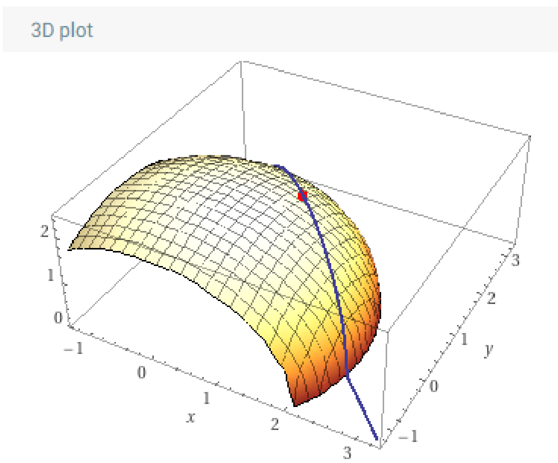
$$\nabla f(x,y) = \lambda \nabla g(x,y)$$

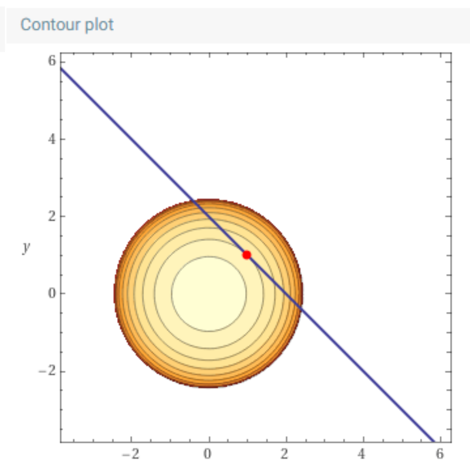
Con lambda encuentra los valores extremos de la función:

Los pasos a seguir son:

1. Define la función de Lagrange (L) $L(x,y,\lambda) = f(x,y) + \lambda g(x,y)$
2. Encuentra las derivadas parciales de L e iguala cero.
3. Soluciona el sistema de ecuaciones de las derivadas parciales. Encuentra puntos críticos.
4. Evalúa los puntos críticos encontrados.
5. Para decidir si es máximo o mínimo aplica el criterio de la segunda derivada.

Problema 1:





16 febrero 2022

Daniel Nuño

Problema 1: resuelve las siguientes problemas de optimización restringida

1) Maximiza $f(x,y)=\sqrt{6-x^2-y^2}$ sujeto a la restricción $g(x,y)=x+y-2=0$ usando multiplicadores de lagrange. Paso 1: define la función L

$$L(x,y,\lambda)=f(x,y)+\lambda g(x,y)$$

$$L(x,y,\lambda)=\sqrt{6-x^2-y^2}+\lambda x+\lambda y-\lambda 2 \quad (1)$$

Paso 2: Derivadas parciales de L

$$\frac{\partial L}{\partial x}=-\frac{x}{\sqrt{-x^2-y^2+6}}+\lambda \quad (2) \qquad \frac{\partial L}{\partial y}=-\frac{y}{\sqrt{-x^2-y^2+6}}+\lambda \quad (3)$$

$$\frac{\partial L}{\partial \lambda}=x+y-2 \quad (4)$$

Paso 3: La derivadas parciales igualar a cero como sistema de ecuaciones utiliza g para sustituir sobre x o y $\quad x=-y+2 \quad (5)$

$$\frac{\partial L}{\partial x}=-\frac{(-y+2)}{\sqrt{(-y+2)^2-y^2+6}}+\lambda=0$$

$$-\frac{-y+2}{\sqrt{-y^2+4y-4-y^2+6}}+\lambda=0$$

$$\frac{y-2}{\sqrt{-2y^2+4y+2}}+\lambda=0$$

multiplica ambos lados por el denominador →

Siendo $\lambda=0$, entonces,

$$y-2=0$$

$$y=2 \quad (6)$$

Sustituye en (4) usando (5)

$$\frac{\partial L}{\partial \lambda}=x+2-2=0=x$$

$x=0$
$y=2$
$\lambda=0$

→

siendo $\lambda = \dfrac{-(y-2)}{\sqrt{-2y^2+4y+2}}$, sustituye en 3 usando tambien $x=-y+2$

$$\frac{\partial L}{\partial y} = \frac{-y}{\sqrt{-2y^2+4y+2}} + \frac{-y+2}{\sqrt{-2y^2+4y+2}} = \frac{-y-y+2}{\sqrt{-2y^2+4y+2}} = 0$$

multiplica ambos lados por el denominador $= -2y+2=0$

$= y = \frac{-2}{-2} = 1$

si $y=1$,

4 $x+y-2=0$ $\quad x=\underline{1}$

$x+1-2=0$ $\quad y=\underline{1}$

$x=1$

Paso 5:

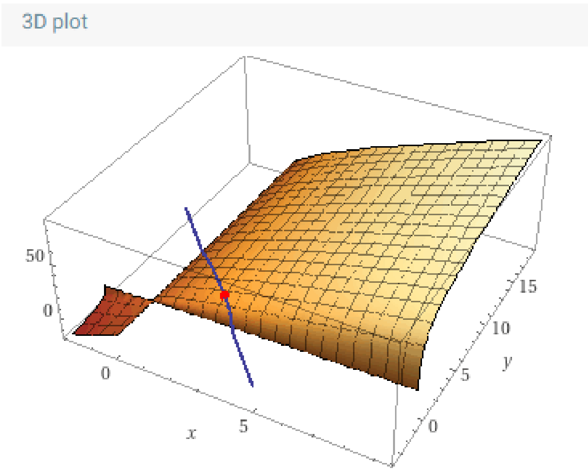
Evalua en $f(1,1)$ y $f(0,2)$ y $f(2,0)$

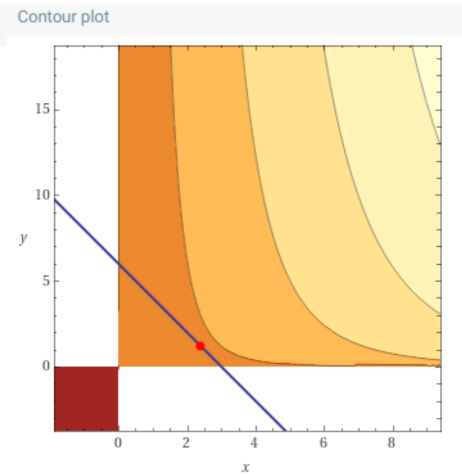
$\underline{f(1,1)=2}$

$f(0,2)=\sqrt{2}$

$f(2,0)=\sqrt{2}$

Problema 2:





2) Máximiza $U(x,y) = 8x^{4/5}y^{1/5}$ sujeto a la restricción $g(x,y) = 4x + 2y = 12$

Paso 1: $L(x,y,\lambda) = 8x^{4/5}y^{1/5} + \lambda 4x + 2y\lambda - 12\lambda$ $= 4x + 2y - 12$ ⑧

Paso 2:

$$\frac{\partial L}{\partial x} = \frac{32\sqrt[5]{y}}{5\sqrt[5]{x}} + 4\lambda = 0 \quad ③$$

$$\frac{\partial L}{\partial y} = \frac{8x^{4/5}}{5y^{4/5}} + 2\lambda = 0 \quad ④$$

$$\frac{\partial L}{\partial \lambda} = 2(2x + y - 6) = 0 \quad ⑤$$

Paso 3: $x = \frac{6-y}{2} = 3 - \frac{y}{2}$ despeje ⑤

$y = 6 - 2x$

en ③ $\frac{32(6-2x)^{1/5}}{5(x)^{1/5}} + 4\lambda = 0$

$-\frac{32(6-2x)^{1/5}}{25(x)^{1/5}} = \lambda$ ⑥

en ④ $\frac{8x^{4/5}}{5(6-2)^{4/5}} + 2\left(-\frac{32(6-2x)^{1/5}}{25(x)^{1/5}}\right) = 0$

$x = \frac{12}{5}$

en ⑤ $y = 6 - 2\left(\frac{12}{5}\right) = \frac{6}{5}$

$x = \frac{12}{5}$

$y = \frac{6}{5}$

Paso 4: $F\left(\frac{12}{5}, \frac{6}{5}\right) = \frac{48(2^{4/5})}{5} \approx 16.71$

Paso 3:

$$\frac{\partial^2 f}{\partial x^2} = -\frac{32\sqrt[5]{y}}{25x^{6/5}} = -\frac{4(2)^{4/5}}{15}$$

$$\frac{\partial^2 f}{\partial y^2} = -\frac{32x^{1/5}}{25y^{9/5}} = -\frac{16(2)^{4/5}}{15}$$

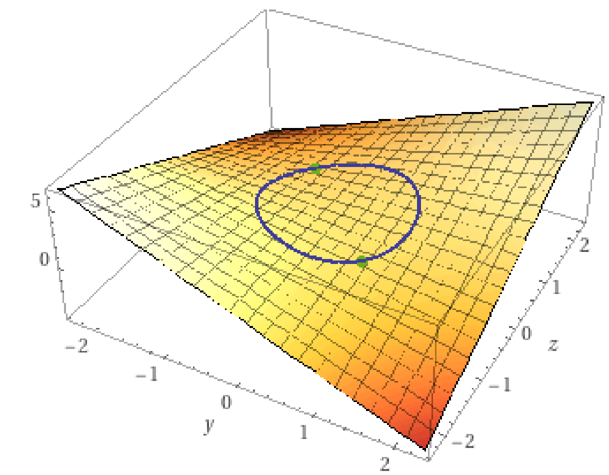
$$\frac{\partial^2 f}{\partial x \partial y} = \frac{32}{25\sqrt[5]{x}\,y^{4/5}} = \frac{8(2)^{4/5}}{15}$$

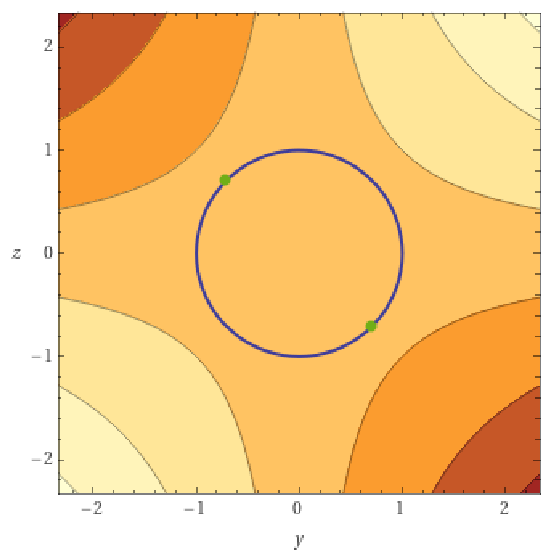
$$\frac{\partial^2 f}{\partial y \partial x} = \frac{32}{25\sqrt[5]{x}\,y^{4/5}} = \frac{8(2)^{4/5}}{15}$$

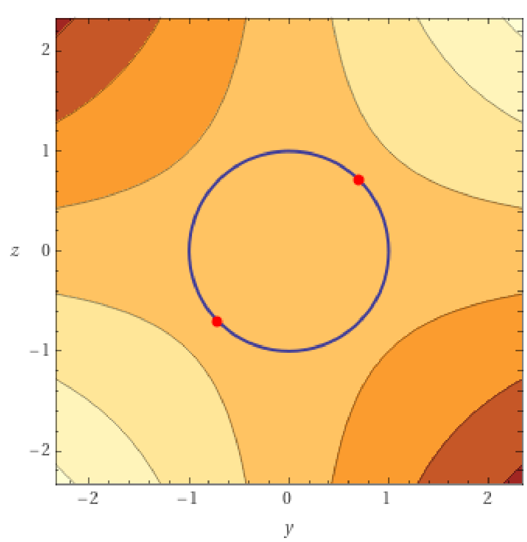
los negativos son más grandes que los positivos y por lo tanto la suma es menor que 0.

$F\left(\frac{12}{5}, \frac{6}{5}\right) \approx 16.71$ es un máximo.

Problema 3:
Graficas







Procedimiento

3. optimiza $f(x,y,z) = yz + xy$ sujeto a ~~constante~~

$g(x,y) = xy = 1 \rightarrow xy - 1 = 0$

$h(y,z) = y^2 + z^2 = 1 \rightarrow y^2 + z^2 - 1 = 0$

como $xy = 1$, entonces

$yz + 1 = f(y,z)$ ①

como $y^2 + z^2 = 1$, entonces $y\ [-1, 1]$, si $y = \pm 1$, $z = 0$

$z = \pm\sqrt{1 - y^2}$ $\quad y = 0$, $z = \pm 1$

Paso 1:

$L(y,z,\lambda) = yz + 1 + \lambda y^2 + \lambda z^2 - 1\lambda$

Paso 2:

$\frac{\partial L}{\partial y} = 2\lambda y + z$ $\qquad \frac{\partial L}{\partial \lambda} = y^2 + z^2 - 1$

$\frac{\partial L}{\partial z} = 2\lambda z + y$

Paso 3:

$\lambda = \frac{-z}{2y}$ $\qquad \lambda = \frac{-y}{2z}$

$y = \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$ $\qquad y = -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$

$z = \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$ $\qquad z = \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$

$f\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = \frac{1}{2}$ Mínimo

$f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{3}{2}$ Máximo

$f\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = \frac{3}{2}$ Máximo

$f\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{2}$ Mínimo

Problema 4:

4: Miniza $f(x,y,z) = x^2 + y^2 + z^2$ cuando $x + y + z = 9$

$x + 2y + 3z = 20$

Usando dos multiplicadores

$\underline{\lambda}$ y $\underline{u}$

Paso 1

$$L(x, y, z, \lambda, u) = f(x,y,z) + \lambda g(x,y,z) + u h(x,y,z)$$

$$= x^2 + y^2 + z^2 + \lambda x + \lambda y + \lambda z + ux + u2y + u3z - 9 - 20$$

Paso 2

$$\frac{\partial L}{\partial x} = \cancel{k} + u + 2x$$

$$\frac{\partial L}{\partial y} = k + u^2 + 2y$$

$$\frac{\partial L}{\partial z} = k + u^3 + 2z$$

$$\frac{\partial L}{\partial \lambda} = x + y + z = 0 \to 9$$

$$\frac{\partial L}{\partial u} = u(3uz + 2y) + x$$

Paso 3:

$x = \frac{7}{2}$,

$y = 3$

$z = \frac{5}{2}$

Paso 4

$f(\frac{7}{2}, 3, \frac{5}{2}) = \frac{55}{2}$ el mínimo

Problema 5:

5.- $xyz = 480\,m^3$

$x = \$5\ m^2$

$y, z = \$3\ m^2$

Paso 1:

① $f(x,y,z) = xyz$

② $xyz = 480$

Porque es un rectángulo entonces

base = TOP $= 5xy$ ó $3xy$

2 lados largos $= 3xz$

2 lados cortos $= 3yz$

⑥ Min $g(x,y,z) = 5xy + 3xy + (3xz)2 + (3yz)2$

$$L(x,y,z) = xyz + \lambda 5xy + \lambda 3xy + 3\lambda xz + 3\lambda yz$$

$$= 480 + \lambda 5xy + 3\lambda xy + 3\lambda xz + 3\lambda yz$$

Paso 2:

$$\frac{\partial L}{\partial x} = 3\lambda(y+z) + 5\lambda y = 0$$

$$\frac{\partial L}{\partial y} = 3\lambda(x+z) + 5\lambda x = 0$$

$$\frac{\partial L}{\partial z} = 3\lambda(x+y) = 0$$

$$\frac{\partial L}{\partial \lambda} = 5xy + 3xy + 3xz + 3yz = 0$$

Paso 3:

$x = 7.11379$

$y = 7.11379$

$z = 9.418505$

Paso 4:

$g(x,y,z) \approx \$1214.54$

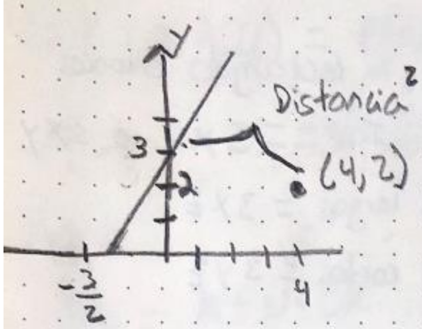
Problema 6:

6. Usa multiplicadores de Lagrange para encontrar el punto sobre la linea $y = 2x + 3$ que esta más cercano al punto $(4,2)$

$f(x) = 2x + 3 = y$

distancia mínima $\sqrt{x^2 + y^2} = \sqrt{(x-4)^2 + (y-2)^2}$

$= \sqrt{(x-4)^2 + (2x+3-2)^2}$



paso 1: usa distancia al cuadrado

$L(x, y, \lambda) = \lambda(2x+3) + (x-4)^2 + (2x+1)^2$

Paso 2:

$\frac{\partial f}{\partial x} = 2(x-4)$

$\frac{\partial f}{\partial y} = 2(y-4)$

$\frac{\partial g}{\partial x} = 2$

$\frac{\partial g}{\partial y} = -1$

Paso 3:

$2(x-4) = 2\lambda$

$x = +4 + 1\lambda$

$x = 4 + \lambda$

$2(y-4) = -1\lambda$

$y = -\frac{1}{2}\lambda + 4$

resolviendo el sistema de ecuaciones usando $\frac{\partial f}{\partial x}$ $\frac{\partial f}{\partial y}$ λ

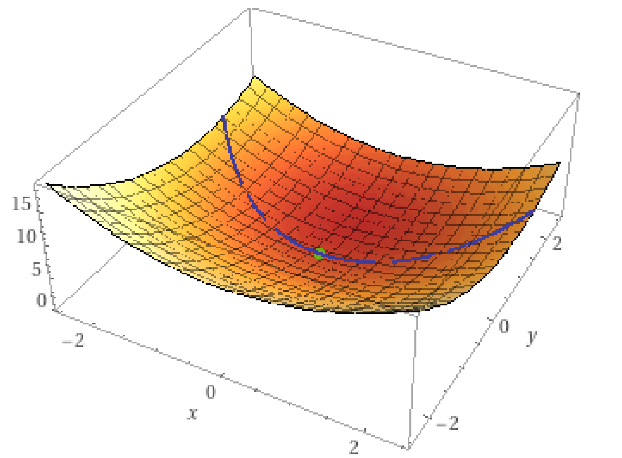
$x = \frac{9}{5}$

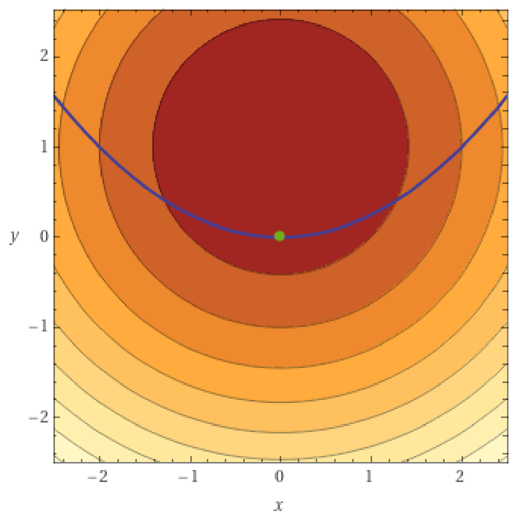
$y = \frac{19}{5}$

$y = \frac{9}{5}(2) - 3 = \frac{19}{5}$. la distancia² es $\frac{81}{5}$

Problema 7

Graficas





Procedimiento

7 usa lagrange para encontrar la distancia minima entre el punto $(0,1)$ y la parabola $x^2=4y$

$$d=\sqrt{(x-0)^2+(y-1)^2}$$

① $d^2 = x^2+(y-1)^2 = x^2+y^2-2y+1$

② $x^2-4y=0$

$$\nabla d^2 = \lambda \nabla (x^2-4y)$$

Paso 2:

$\frac{\partial d^2}{\partial x} = 2x$ $\qquad$ $\frac{\partial g}{\partial x} = 2x$

$\frac{\partial d^2}{\partial y} = 2(y-1)$ $\qquad$ $\frac{\partial g}{\partial y} = -4$

paso 3:

$2x = 2x\lambda$ $\qquad$ $2(y-1) = -4\lambda$ $\qquad$ $x^2 = 4(-1)$

$1=\lambda$ $\qquad$ $y = -2\lambda+1$ $\qquad$ $x^2=-4$

$y = -2+1$ $\qquad$ $x=\sqrt{-4}$

$x = \pm 2$

Paso 4:

ó si $\lambda = 0$ entonces:

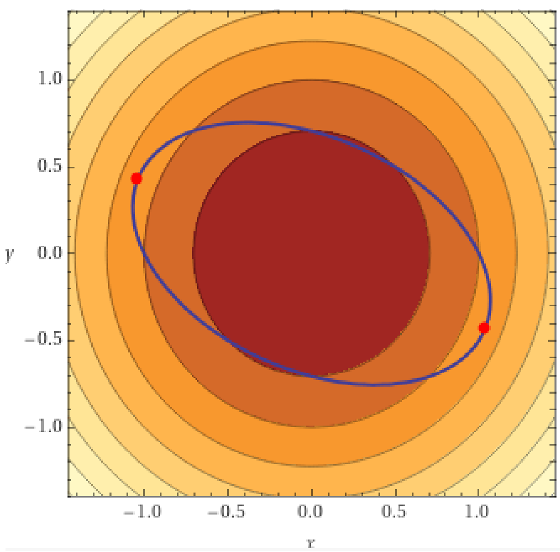
$x=0$

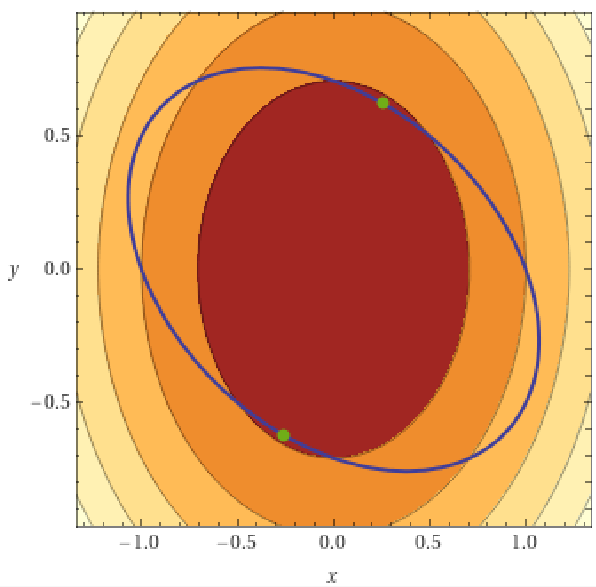
$y=0$

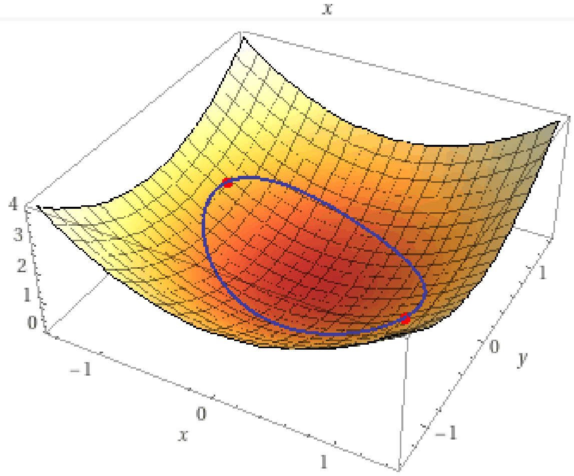
Paso 5:

$0^2-(0-1)^2 = 1$ es minimo de la distancia

Problema 8:
Graficas







8. Usa multiplicadores de lagrange para encontrar el mínimo y máximo de las distancias entre el elipse $x^2 + xy + 2y^2 = 1$ y el origen

Paso 1:

① $d^2 = x^2 + y^2 = f$

② $x^2 + xy + 2y^2 - 1 = 0 = g$

Paso 2:

③ $\frac{\partial f}{\partial x} = 2x$ ⑤ $\frac{\partial g}{\partial x} = 2x + y$

④ $\frac{\partial f}{\partial y} = 2y$ ⑥ $\frac{\partial g}{\partial y} = x + 4y$

Paso 3:

$2x = 2x\lambda + \lambda y$

$\frac{2x}{2x+y} = \lambda$

$\frac{1}{y} = \lambda$

$2y = \lambda(x - 4y)$

$2y = \frac{1}{y}(x - 4y)$

$2y^2 + 4y = x$

Sustituyendo los valores de λ en la ecuación ⑤ y ⑥

los puntos críticos son

$(x, y) = -1.03, 0.42 \approx 1.2612 \rightarrow$ máximo

$(x, y) = 1.03, -0.42 \approx 1.2612 \rightarrow$ máximo

$(x, y) = -0.25, -0.62 \approx 0.45 \rightarrow$ mínimo

$(x, y) = 0.25, 0.62 \approx 0.45 \rightarrow$ mínimo

Problema 2:

Problem 2: Using Transformed Features

Read and reproduce the example about the Boston housing dataset given in Gplearn: Symbolic Transformer. Then, explain how the *symbolic transformer* method helps to improve the regression's performance. Upload your results to Github in the form of a Jupyter notebook, then make it interactive using Binder, hence submit your results through both links. The use of Google Colab is highly recommended.

- Liga a github:
- Liga usando Binder

Problema 3:

Problem 3: LS-SVM: Regression

Consider the following optimization problem:

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

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

problema 3 de la tarea — 16 feb ,22

$$\min_{w,b,e} \rho(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{n} e_k^2$$

sujeto $y_k = w^T \varphi(x_k) + b + e_k$, $k = 1, \dots, N$

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

KKT1: primer orden

(1) $\nabla_w L = w - \sum \alpha_k \varphi(x_k) = 0 \Rightarrow w \sum \alpha_k \varphi(x_k)$ — w es un vector

(2) $\frac{\partial L}{\partial b} = -\sum_{k=1}^{N} \alpha_k = 0 \Rightarrow \sum_{k=1}^{N} \alpha_k = 0$

(3) $\nabla \frac{\partial L}{\partial e_k} = \gamma e_k - \alpha_k = 0$ $\quad k = 1, \dots, N$ → es un vector

KK2:

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

(1) y (3) en (4)

$$\sum_{i=1}^{N} \alpha_i \varphi^T(x_i) \varphi(x_k) + b + \frac{\alpha_k}{\gamma} = y_k \qquad (5)$$

Ahora desarrollando la ecuación 5 en $K=1,2,\dots,N$

Para $K=1$

$$\sum_{i=1}^{N} \alpha_i \varphi^T(X_i)\varphi(X_1) + b + \frac{\alpha_1}{\Gamma} = Y_1$$

Para $K=2$

$$\sum_{i=1}^{N} \alpha_i \varphi^T(X_i)\varphi(X_2) + b + \frac{\alpha_2}{\Gamma} = Y_2$$

Para $K=N$

$$\sum_{i=1}^{N} \alpha_i \varphi^T(X_i)\varphi(X_N) + b + \frac{\alpha_N}{\Gamma} = Y_N$$

Ahora para cada K, desarrolla la sumatoria de i

Para $K=1$, $i=1,2,\dots,N$.

$$\alpha_1 \varphi^T(X_1)\varphi(X_1) + \alpha_2 \varphi^T(X_2)\varphi(X_1) + \dots + \alpha_N \varphi^T(X_N)\varphi(X_1) + b + \frac{\alpha_1}{\Gamma} = Y_1$$

Para $K=2$, $i=1,2,\dots,N$

$$\alpha_1 \varphi^T(X_1)\varphi(X_2) + \alpha_2 \varphi^T(X_2)\varphi(X_2) + \dots + \alpha_N \varphi^T(X_N)\varphi(X_2) + b + \frac{\alpha_2}{\Gamma} = Y_2$$

Para $K=N$, $i=1,2,\dots,N$

$$\alpha_1 \varphi^T(X_1)\varphi(X_N) + \alpha_2 \varphi^T(X_2)\varphi(X_N) + \dots + \alpha_N \varphi^T(X_N)\varphi(X_N) + b + \frac{\alpha_N}{\Gamma} = Y_N$$

El desarrollo anterior es una matriz que se puede factorizar en una matriz y 3 vectores.

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

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

$$\underline{K}\alpha + b1_v + \frac{1}{r}\alpha = Y \Rightarrow (\underline{K} + 1/r\, I)\alpha + b\,1_v = Y$$

$$\sum_{k=1}^{N} \alpha_k = 0, \quad \alpha^T 1_v = 0, \quad 1_v^T \alpha = 0$$

A final resumimos el desarrollo y el sistema de ecuaciones como matriz multiplicando por un vector

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