Optimización Tarea 3

Francisco Javier Peralta Ramírez

1. Escribe la expansión de Taylor de segundo orden para

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \qquad \boldsymbol{x} in \mathbb{R}^2$$

   a) La expansión de Taylor de segundo orden para variables multidimensionales se escribe como

$$f(\boldsymbol{x}) \approx f(\boldsymbol{a}) + Df(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + \frac{1}{2}Hf(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a})$$

   Donde $Hf(\boldsymbol{a}) = DDf(\boldsymbol{a})$. es decir la *matriz Hessiana*.

   b) En el caso de nuestra función, podemos calcular la primera derivada y el Hessiano

$$Df(\boldsymbol{x}) = \begin{pmatrix} 400(x_2 - x_1^2)x_1 + 2(1 - x_1) & 200(x_2 - x_1^2) \end{pmatrix}$$

$$Hf(\boldsymbol{x}) = \begin{pmatrix} 400x_2 - 600x_1^2 - 2 & 400x_1 \\ 400x_1 & 200 \end{pmatrix}$$

   Ahora podemos substituir en la formula anterior

$$f(\boldsymbol{x}) \approx f(\boldsymbol{a}) + \begin{pmatrix} 400(a_2 - a_1^2)a_1 + 2(1 - a_1) \\ 200(a_2 - a_1^2) \end{pmatrix}^T \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 400x_2 - 600x_1^2 - 2 & 400x_1 \\ 400x_1 & 200 \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ x_2 - a_2 \end{pmatrix}$$

2. Supngamos que observamos m valores de una función g en los puntos $x_1, x_2, \cdots, x_m$ por lo que conocemos los valores de $g(x_1), g(x_2), \cdots, g(x_m)$. Queremos aproximar la función $g(\cdot) : \mathbb{R} \to \mathbb{R}$ por un polinomio

$$h_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

   con $n < m$

   Para aproximar la función queremos minimizar el error dado por

$$E = \sum_{i=1}^{m} (g(x_i) - h(x_i))^2$$

   Lo caul se puede hacer resolviendo el sistema de ecuaciones dado por $\mathcal{D}E = \mathbf{0}$ donde derivamos E con respecto a $[a_1, a_2, \cdots, a_n]$

   a) Genera observaciones del modelo

$$g(x) = \frac{sin(x)}{x} + \eta \qquad x \in [0.1, 10]$$

   Donde $\eta \sim \mathcal{N}(0, 0.1) 0.1 = x_1 < x_2 < \cdots < x_m = 10.$

   b) Aproxima $g(\cdot)$ por $h_n(x)$ para $n = 2, \cdots, 5$

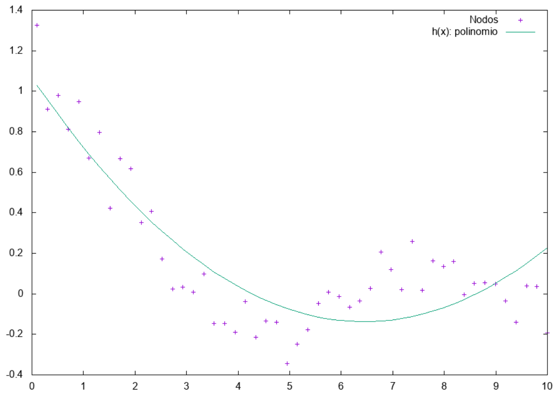
$$h_2(x) = 1.067 - 0.373x + 0.029x^2$$

$$h_3(x) = 1.417 - 0.789x + 0.131x^2 - 0.006x^3$$

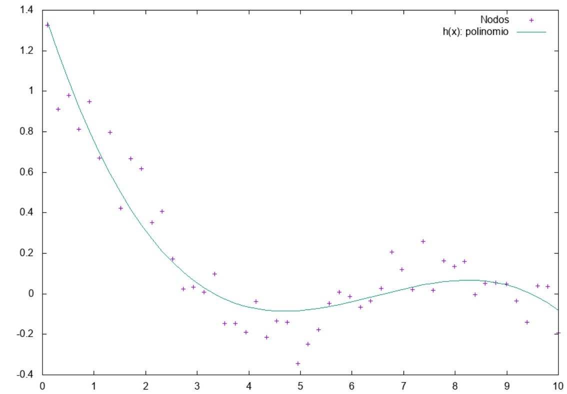
$$h_4(x) = 1.243 - 0.446x - 0.020x^2 + 0.016x^3 - 0.001165x^4$$

$$h_5(x) = 1.114 - 0.066x - 0.283x^2 + 0.086x^3 - 0.008886x^4 + 0.0003x^5$$

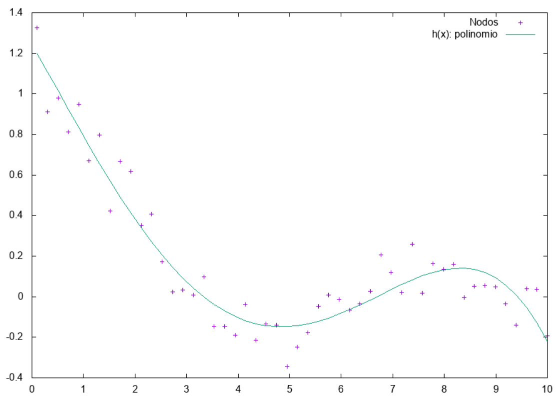
c) Grafica en la misma gráfica $h_n(x)$ para $n = 2, \cdots, 5$ y $x_i, g(x_i), i = 1, 2, \cdots, m$



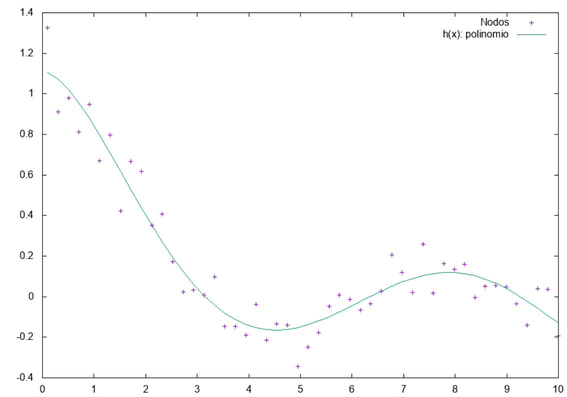
(a) Polinomio grado 2



(b) Polinomio grado 3



(a) Polinomio grado 4



(b) Polinomio grado 5

3. Dada una $f(x)$ continua en $[a, b]$ encuentra el polinomio de aproximación de grado n

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

que minimiza

$$\int_a^b [f(x) - p(x)]^2 dx$$

Al igual que en el ejercicio anterior. Calculasmos las derivadas parciales con respecto a $[a_0, a_1, \cdots, a_n]$ y nos queda un sistema de ecuaciones.

$$\frac{\delta}{\delta a_0} = 2 \int_a^b f(x) - p(x) dx$$

$$\frac{\delta}{\delta a_i} = 2 \int_a^b x^i f(x) - x^i p(x) dx$$

Podemos igualar a cero y generar el sistema de ecuaciones:

$$\begin{pmatrix} \int_a^b 1 & \int_a^b x & \cdots & \int_a^b x^n dx \\ \int_a^b x & \int_a^b x^2 & \cdots & \int_a^b x^n dx \\ \vdots & \ddots & \cdots & \vdots \\ \int_a^b x^n & \int_a^b x^{n+1} & \cdots & \int_a^b x^2 n dx \end{pmatrix} \times \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \int_a^b f(x) \\ \int_a^b x f(x) \\ \vdots \\ \int_a^b x^n f(x) \end{pmatrix}$$

4. Para la distribución normal $\mathcal{N}(\mu, \sigma^2)$ que tiene una función de densidad de probabilidad

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

La función de desindad de probabilidad correspondiente para una muestra $\{x_i\}_{i=1}^n$ de n variables independientes e identicamente distribuidas es (verosimilitud):

$$\mathcal{L}(\mu,\sigma^2;x) = f(x_1,x_2,\cdots,x_n|\mu,\sigma^2) = \prod_{i=1}^n f(x_i|\mu,\sigma^2)$$

Con $x = x_1, x_2, \cdots, x_n$. Calcula

$$(\mu^*,\sigma^*) = arg\max_{\mu,\sigma} \mathcal{L}(\mu,\sigma^2;x)$$

$$arg\max_{\mu,\sigma} \mathcal{L}(\mu,\sigma^2;x) = (2\pi\sigma^2)^{(-n/2)} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right)$$

Podemos ver que el problema de maxima verosimilitud es el mismo que el de log verosimilitud. Por lo que el problema se puede ver como

$$arg\max_{\mu,\sigma} l(\mu,\sigma^2;x) = -\frac{n}{2}ln(2\pi) + \frac{n}{2}\sigma^2 + -\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2$$

Para encontrar el máximo, derivamos con respecto a μ y σ^2 e igualamos a 0

$$\frac{\delta}{\delta\mu} = \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu) = 0$$

Esto sólo es cero cuando

$$\sum_{i=1}^n (x_i-\mu) = 0$$

$$\left(\sum_{i=1}^n x_i\right) - n\mu = 0$$

$$\mu = \frac{1}{n}\sum_{i=1}^n x_i$$

Lo cual equivale al promedio muestral. Y para σ^2

$$\frac{\delta}{\delta\sigma^2} = -\frac{n}{2\sigma^2} + \left[\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^2\right]\left(-\frac{1}{(\sigma^2)^2}\right) = 0$$

$$= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 - n\right] = 0$$

El cual es cero sólo cuando

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^n (x_i-\mu)^2$$

Lo cual equivale a la varianza muestral

5. *a*) Suponga que $f : \mathbb{R} \to \mathbb{R}$ es convexa y $a, b \in dom f$ con $a < b$. Muestra que

$$a(f(x) - f(b)) + x(f(b) - f(a)) + b(f(a) - f(x)) \geq 0 \qquad \forall x \in [a, b]$$

Podemos reescribir la formula como

$$f(x)(a - b) + f(b)(x - a) + f(a)(b - x) \geq 0$$

También podemos expresar x como $x = \alpha(a) + (1 - \alpha)b$. Substituimos en nuestra ecuación.

$$f(\alpha(a) + (1 - \alpha)b)(a - b) + f(b)(\alpha(a) + (1 - \alpha)b - a) + f(a)(b - \alpha(a) - (1 - \alpha)b) \geq 0$$

$$f(\alpha(a) + (1 - \alpha)b)(a - b) + f(b)(-a(1 - \alpha) + (1 - \alpha)b) + f(a)(-\alpha(a) - \alpha(b)) \geq 0$$

$$f(\alpha(a) + (1 - \alpha)b)(a - b) + f(b)((1 - \alpha)(b - a)) \geq \alpha f(a)(a - b)$$
$$f(\alpha(a) + (1 - \alpha)b)(a - b) \geq \alpha f(a)(a - b) + f(b)((1 - \alpha)(a - b))$$

Por definición, una función es convexa cuando

$$f[\alpha x_1 + (1 - \alpha)x_2] \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$$

Como $(a - b) \leq 0$ se cumple que es convexa

b) Suponga que $f : \mathbb{R}^n \to \mathbb{R}$ es convexa con $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ y $\boldsymbol{b} \in \mathbb{R}^n$. Muestra que la función $g : \mathbb{R}^m \to \mathbb{R}$ definida por $g(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$, con $dom\, g = \{x | \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \in dom\, f\}$, es convexa.
Tomemos una $\boldsymbol{z} = \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}$

$$g(\boldsymbol{z}) = g(\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) = f(\alpha\boldsymbol{A}\boldsymbol{x} + (1 - \alpha)\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})$$

$$g(\boldsymbol{z}) = g(\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) = f(\alpha\boldsymbol{A}\boldsymbol{x} + (1 - \alpha)\boldsymbol{A}\boldsymbol{y} + (\alpha + (1 - \alpha))\boldsymbol{b})$$

$$g(\boldsymbol{z}) = g(\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) = f(\alpha(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) + (1 - \alpha)(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}))$$

Consideremos $x_2 = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ y $y_2 = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$

$$g(\boldsymbol{z}) = f(\alpha x_2 + (1 - \alpha)y_2)$$

Como f es convexa

$$g(\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) \leq \alpha f(x_2) + (1 - \alpha)f(y_2)$$

Substituimos valores

$$g(\alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) + (1 - \alpha)f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})$$

Por lo que cumple el criterio de ser convexa