Tema 3.1 : Regresión polinomial (Univariable)

En este apartado se provee una descripción del proceso de normalización. Ademas se muestra el resultado de un ejemplo de regresión haciendo uso de este proceso. De esta lectura se obtendrá una comprensión de la importancia de la normalización en aplicaciones de regresión.

Considérese una hipótesis que es un polinomio de grado g:

$$y(x) = \omega_0 x^0 + \omega_1 x^1 + \omega_2 x^2 + ... + \omega_g x^g \tag{1}$$

Existen diferentes comportamientos de esta expresión dependiendo el valor de x. Estos comportamientos pueden ser clasificados en dos casos: El primero es cuando x se encuentra en los rangos $[x < -1] \cup [x > 1]$ y el segundo es cuando $[-1 < x < 1]$.

En el primer caso, el valor del j-esimo termino ($\forall j \in [0, g]$) del polinomio tendera a ser mas grande a medida que j es mayor. Por ejemplo, cuando $x = 5$, los valores de x^j serán: $1, 5, 25, 125, 625, 3125, \ldots$. Esto implica que los términos donde x se eleva a una mayor potencia estarán aportando mas al valor de y. Este comportamiento se ve reflejado en la derivada de la función de costo con respecto al coeficiente ω_j. Considérese el MSE como función de costo, la derivada resulta como:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (y_i(x_i) - T_i) x_i^j \tag{2}$$

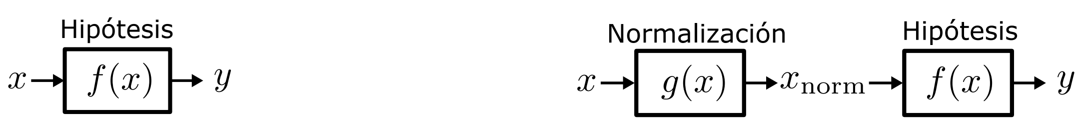
por lo cual, el termino $(y_i(x_i) - T_i)$ es multiplicado por x_i^j que provocará un valor de derivada cuya magnitud crecerá exponencialmente a conformen j es mayor. Debido a que el factor de aprendizaje α es constante e igual para la actualización de cualquier coeficiente:

$$\omega_{j,t+1} = \omega_{j,t} - \alpha \frac{\partial C(\mathbf{w})}{\partial \omega_{j,t}} \tag{3}$$

los coeficientes pertenecientes a los términos donde x se eleva a potencias mas grandes se actualizaran mas agresivamente, es decir, el nuevo valor del coeficiente será en cada iteración muy distinto al anterior. En cambio los coeficientes asociados a términos donde x se eleva a potencias bajas, se actualizaran mas suavemente.

En el segundo caso ($[-1 < x < 1]$), se tiene el escenario opuesto. conforme x se eleva a potencias mayores el resultado es un valor cada vez mas pequeño, causando que los términos donde x se eleva a potencias mayores contribuyen en menor medida al resultado y. Asimismo, la magnitud de la derivada a medida que j crece, tiende a ser menor, por o que la actualización de ω sera casi inexistente.

Idealmente, todos los coeficientes ω deben actualizarse en la misma proporción en cada iteración. Para acercarse a este hecho se utiliza la *normalización*. El proceso representado en la Figura 1a corresponde al ejercicio de regresión descrito hasta el momento. Sin embargo, es necesario agregar un pre-proceso adicional, el cual consiste en aplicar una función $g(x)$ que mapee de manera linear los valores de x a un conjunto de nuevos valores (x_{norm}) cuyo rango sea definido por el diseñador (Figura 1b).



(a) (b)

Figure 1: Diagrama de bloques de a) el proceso de calculo de y y b) el proceso de calculo de y empleando muestras normalizadas

Para determinar la ecuación $g(x)$ considere la Figura 2. Si $x \in [R1, R2]$ y se desea mapear sus valores de manera lineal a un nuevo rango $[S1, S2]$ es necesario obtener la ecuacion de una recta $z = g(x) = ax + b$, donde a representa la pendiente y b el valor de $g(x)$ cuando $x = 0$.

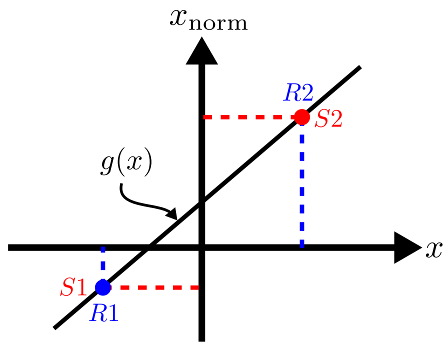


Figure 2: Representación de la función $g(x)$ para el proceso de normalización. (La ubicación de los valores $R1, R2, S1, S2$ es arbitraria en la ilustración, estos pueden tener diferente ubicación cada uno sin afectar el resultado).

La pendiente a es obtenida como:

$$a = \frac{S2 - S1}{R2 - R1} \tag{4}$$

mientras que el valor b puede ser expresado a partir de los valores conocidos de salida de la función $g(x)$. Por ejemplo, se sabe que $S2 = g(R2)$ por tanto:

$$S2 = ax + b = \left(\frac{S2 - S1}{R2 - R1}\right) R2 + b \tag{5}$$

entonces:

$$b = S2 - \left(\frac{S2 - S1}{R2 - R1}\right) R2 \tag{6}$$

Así, $g(x)$ queda como:

$$z = g(x) = \left(\frac{S2 - S1}{R2 - R1}\right) x + \left(S2 - \left(\frac{S2 - S1}{R2 - R1}\right) R2\right) \tag{7}$$

$$= \boxed{\left(\frac{S2 - S1}{R2 - R1}\right)(x - R2) + S2} \tag{8}$$

Ahora que se tiene la función para normalizar, se toma como ejemplo en conjunto de muestras de la Figura 3 y sobre estas se realiza este preprocesamiento a los valores de x resultando ahora en el ajuste mostrado en la Figura 4a. Asimismo, se muestra en la Figura 4b el costo obtenido en cada iteración. Para lograr este resultado se utilizo $\alpha = 0.75$, $S1 = -1$, $S2 = 1$ y un número de iteraciones igual a 50.

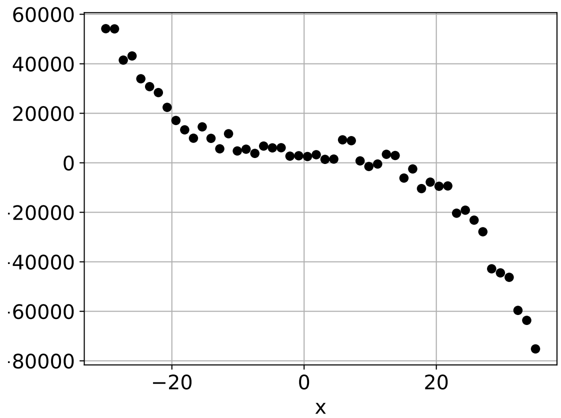
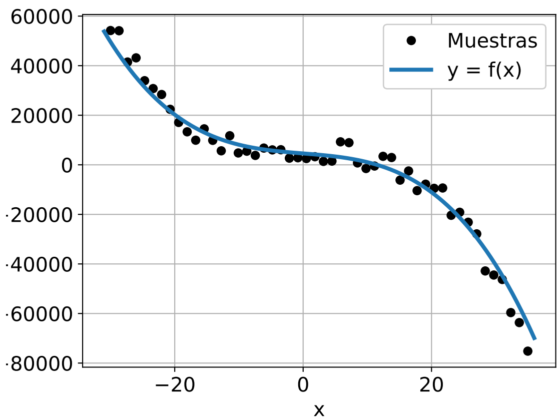
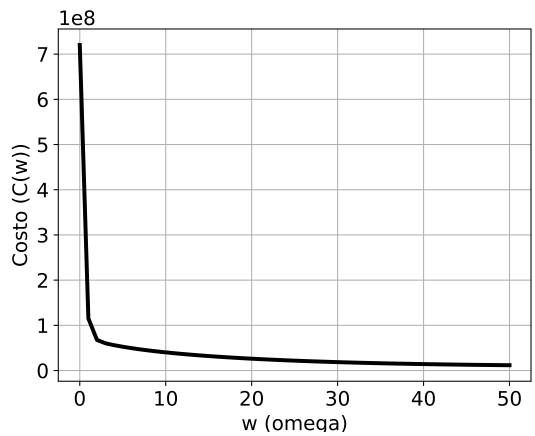


Figure 3



(a)



(b)

Figure 4: Ejemplo de un conjunto de veinte muestras tomado como referencia para un problema de regresión polinomial a) Ajuste logrado con después de haber incorporado el proceso de normalización a los valores x. b) Costo obtenido en cada iteración del algoritmo de búsqueda.

Ejercicios

1. Bajo el paradigma de regresión polinómica, realizar en algoritmo de entrenamiento para lograr el ajuste de diferentes hipótesis a los datos de los conjuntos de entrenamiento generados por las siguientes funciones:

   (a) $y = sin(x)$ para $x = [0, 2\pi]$ (10 puntos para x)

   (b) $y = 5 + 7\log(x)$ para $x = [0, 100]$ (100 puntos para x)

   (c) $y = \frac{1}{(1+e^{-x})}$ para $x = [-5, 5]$ (100 puntos para x)

   Para cada uno de los incisos listados anteriormente, se deben probar al menos tres hipótesis distintas (tres grados diferentes), por ejemplo:

(a) $y(x) = \omega_0 x^0 + \omega_1 x^1 + \omega_2 x^2$

(b) $y(x) = \omega_0 x^0 + \omega_1 x^1 + \omega_2 x^2 + \omega_3 x^3$

(c) $y(x) = \omega_0 x^0 + \omega_1 x^1 + \omega_2 x^2 + \omega_3 x^3 + \omega_4 x^4$

Ademas se deberá explorar el entrenar con datos normalizados y sin normalizar. Para cada implementación (Seis implementaciones correspondientes a los tres incisos empleando x sin normalizar y normalizado) se deben generar al menos las siguientes dos graficas:

(a) Graficar T vs x y sobre la misma grafica $y(x)$ vs x (Mostrar la predicción de la hipótesis para al menos un 50% adicional al rango de x)

(b) Grafica costo vs Iteraciones