Machine Learning

Programa de Ingeniería de Sistemas

Tema: Regresión Lineal – Gradiente Descendente

LuzStellaRoblesPedrozo

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error se plantea el siguiente problema de optimización:

$$\min_\theta \; J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2.$$

Para resolverlo, se puede optar por:

- El método Analítico, haciendo uso de Ecuaciones Normales
- El método del Gradiente Descendente

$J(\theta)$ θ's $h(x^{(i)})$'s $y^{(i)}$' s

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**
  - El método del Gradiente Descendente es un algoritmo de optimización que utiliza el valor del gradiente de la función a optimizar para dirigirse hacia un mínimo local.
  - Es un método iterativo de primer orden (que contiene solo derivadas simples).
  - Si la función a optimizar es convexa, este método siempre convergerá al óptimo local dada una correcta escogencia de parámetros.
  - La intuición tras este método es que siendo el gradiente, el vector que apunta en la dirección de mayor crecimiento de la función a optimizar, es lógico entonces moverse en la dirección contraria a este, para alcanzar un óptimo local.

Regresión Lineal

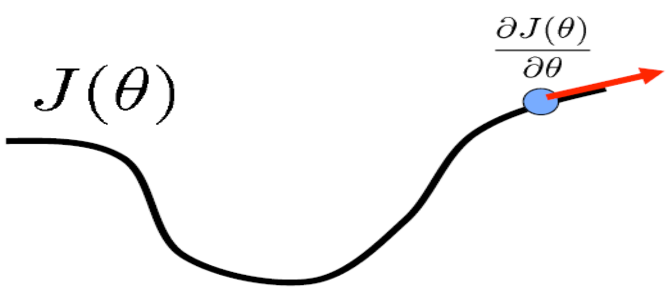
Aprendizaje

- **El método del Gradiente Descendente**

Fórmula general:

$$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

Captura la intuición de moverse en al dirección contraria al gradiente, al anteponer un signo negativo a este.



Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

  Para poder aplicarlo, necesitamos conocer el gradiente de la función a optimizar

$$\left(\frac{\partial J(\theta)}{\partial \theta}\right)$$

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

Entonces, tenemos que la función de error en la Regresión Lineal es:

$$J(\theta) = \frac{1}{2N} \sum_{n=1}^{N} (h(x_n) - y_n)^2$$

Nótese que ahora la sumatoria está multiplicada por $\frac{1}{2N}$. Este es un pequeño artificio para facilitar la derivación de la regla de actualización de Gradiente Descendente, y no afecta el resultado.

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

El gradiente entonces es:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{N}\sum_{n=1}^{N}(h(x_n) - y_n)x_{nj}$$

j	es el subíndice de los θ's
n	es el subíndice de la entrada en la matriz X (el registro n en la matriz X)
X_{nj}	es un valor en matriz X ubicado en la fila n, columna j (fila n de N registros, columna j de k parámetros: θ_0 , θ_1 , θ_2 ... θ_k

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

Por lo anterior, la regla de actualización de Gradiente Descendente sería:

$$\theta_{\boldsymbol{j}} = \theta_j - \alpha \frac{1}{N} \sum_{n=1}^{N} (h(X_n) - Y_n) X_{nj}$$

Dónde:

$h(X_n)$ en Regresión Lineal es $\theta^{\mathcal{T}} \mathcal{X}$

$h(X_n)$ en Regresión Logística es $\dfrac{1}{1 + e^{-\theta^t x}}$

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

  Teniendo la regla, el algoritmo sería el siguiente:

  1. Hacer, para toda j
  2. La función…

  $$\theta_{\boldsymbol{j}} = \theta_j - \alpha \frac{1}{N} \sum_{n=1}^{N} (h(X_n) - Y_n) X_{nj}$$

  3. Mientras que $\alpha \left|\left| \frac{\partial J(\theta)}{\partial \theta_j} \right|\right| > \epsilon$

  Nota: La actualización de los θ_j debe ser simultánea (primero se calculan todos los θ_j y luego se reemplazan los anteriores).

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

Mientras que $\alpha || \frac{\partial J(\theta)}{\partial \theta_j} || > \epsilon$

$\alpha || \frac{\partial J(\theta)}{\partial \theta_j} || > \epsilon$ = 0.001 (por ejemplo)

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

La regla de actualización del Gradiente Descendente sería:

$$\theta_j = \theta_j - \alpha \frac{1}{N} \sum_{n=1}^{N} \left[\left(\frac{1}{1 + e^{-\theta^T x_n}} - y_n \right) X_{nj} \right]$$

El vector $\vec{\theta}$ es igual a $[\theta_0 , \theta_1 , \theta_2]$

El vector $\vec{\theta}$ inicialmente tiene los valores en cero: [0, 0, 0]

El subíndice j representa la cantidad de θ's

α vale 0.1 (sugerencia)

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

  Cálculo de θ_0 :

$$\vec{\theta} = [\theta_0 , \; \theta_1 , \; \theta_2]$$

Suponiendo que se cuenta con una matriz X, con m columnas (una por cada θ)

$$X = \begin{array}{c} \overbrace{\hphantom{1 \quad x_{n1} \quad x_{n2}}}^{m} \\ \begin{pmatrix} 1 & & \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \\ \vdots & \vdots & \vdots \\ 1 & & \end{pmatrix} \end{array} \qquad \begin{array}{c} Y \\ \begin{pmatrix} \\ \vdots \\ y_n \\ \vdots \\ \\ \end{pmatrix} \end{array} \left.\vphantom{\begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}}\right\} N \text{ registros}$$

x_{no}

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

Primera interacción de θ_0 :

$$\theta_0 = \theta_0 - \alpha \frac{1}{N} \sum_{n=1}^{N} \left[\left(\frac{1}{1 + e^{-\theta^T x_n}} - y_n \right) X_{n0} \right]$$

En la sumatoria, en el término $1 + e^{-\theta^T x_n}$ es conveniente:

- Realizar la operación $-\theta^T x_n$ dentro del ciclo, pero justo antes de hacer la operación $\frac{1}{1+ e^{-\theta^T x_n}}$ cada vez!
- Para este primer caso sería así:

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

- Para este primer caso sería así:
  Vector θ^T * la fila 1 de la matriz X (n = 1).

$$\theta^T x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3*1} * x_1$$

$$\theta^T x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3*1} * \begin{bmatrix} x_{10} & x_{11} & x_{12} \end{bmatrix}_{1*3}$$

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

- Para este primer caso sería asÍ:
  Vector θ^T * la fila 1 de la matriz X (n = 1).

$$\theta^T x_1 = \theta_0 x_{10} + \theta_1 x_{11} + \theta_2 x_{12}$$

$$\theta^T x_1 = 1 + \theta_1 x_{11} + \theta_2 x_{12}$$

$\theta^T x_1 =$ da un número (que llamaremos s1)

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

- Para este segundo caso sería así:
  Vector θ^T * la fila 2 de la matriz X (n = 1).

$$\theta^T x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3*1} * \; x_2$$

$$\theta^T x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3*1} * \begin{bmatrix} x_{20} & x_{21} & x_{22} \end{bmatrix}_{1*3}$$

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

- Para este segundo caso sería asÍ:
  Vector θ^T * la fila 2 de la matriz X (n = 2).

$$\theta^T x_2 = \theta_0 x_{20} + \theta_1 x_{21} + \theta_2 x_{22}$$

$$\theta^T x_2 = 1 + \theta_1 x_{21} + \theta_2 x_{22}$$

$\theta^T x_2 =$ da un número (que llamaremos s1)

Etc., hasta llegar a N registros.

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

- Al valor obtenido al evaluar la expresión $\frac{1}{1+e^{-\theta^T x_1}}$, se le resta el valor almacenado en el vector $\vec{Y}$, en la fila 1, de N-filas en la matriz, así:

$$\frac{1}{1+e^{-\theta^T x_1}} - Y_1 \qquad \Bigg| \qquad \frac{1}{1+e^{-s1}} - Y_1$$

- El resultado de la operación anterior, es multiplicado por su respectiva X_{n0}, que en este caso sería:

$$X_{n0} = X_{10}$$

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

$$X_{n0} = X_{10}$$

Es uno (1) porque se está evaluando ahora mismo la fila 1 de la matriz X.

Es cero (0) en la posición, porque se está intentando calcular el valor de θ_0

En este caso, el valor de X_{10} es igual a 1. (Columna insertada en la primera columna de la matriz X.

Regresión Lineal

Aprendizaje

El método del Gradiente Descendente

Una vez se obtenga el cálculo de la sumatoria, este resultado es multiplicado por $\alpha \frac{1}{N}$.

El resultado de toda la operación debe ser guardado en una variable porque esta variable es utilizada para controlar el ciclo (para el cálculo del valor de cada θ).

$$\alpha || \frac{\partial J(\theta)}{\partial \theta_j} || > \epsilon$$

Nota: La actualización de los θ_j debe ser simultánea (primero se calculan todos los θ_j y luego se reemplazan los anteriores).

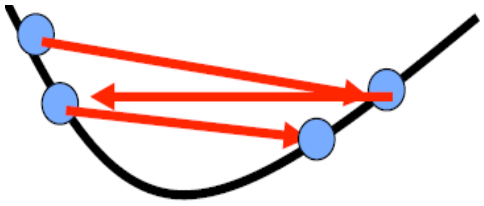
Regresión Lineal

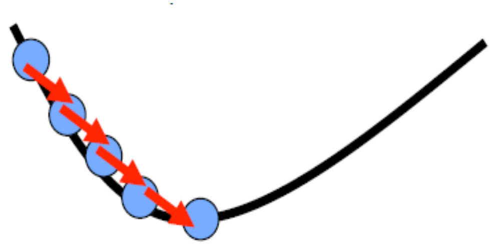
Aprendizaje

- **El método del Gradiente Descendente**

  α se conoce como la tasa de aprendizaje (el tamaño de los pasos que da el algoritmo), y debe ser escogido cuidadosamente puesto que:

  - Un valor muy alto puede hacer que el algoritmo diverja
  - Un valor muy pequeño puede demorar demasiado tiempo en converger.





Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**
  - Al usar Gradiente Descendente para entrenar un modelo de Regresión Lineal, es conveniente escalar y normalizar las entradas.
  - Escalando y normalizando usualmente acelera el proceso de aprendizaje, al poder usarse una tasa de aprendizaje mayor.
  - Una forma de normalizar la media de los datos, que significa llevar $\mu = 0$, y llevarlos aproximadamente al intervalo [-1, 1], es la siguiente:

$$x_i = \frac{x_i - \mu_i}{max_i - min_i}$$

para cada variable de entrada x.

Regresión Lineal

Aprendizaje

- **El método del Gradiente Descendente**

  Ejemplo de normalización de datos:

$$x_i = \frac{x_i - \mu_i}{max_i - min_i}$$

$\mathcal{X}$ = Edad [20, 30, 40, 50, 60]

$$x_1 = \frac{\mathbf{20} - 40}{60 - 20} = \frac{-20}{40} = \frac{-1}{2}$$

$$x_2 = \frac{\mathbf{30} - 40}{60 - 20} = \frac{-10}{40} = \frac{-1}{4}$$

Actividad Extra-clase

Taller

No aplica aún!

Referencias

Material de apoyo de la semana registrados en SAVIO.

Gracias!