Regresión Lineal

Grupo Análisis Numérico y Financiero

Universidad de Antioquia

28 de agosto de 2015

Outline

Next Subsection

Regresión lineal con una variable

Componentes

Se miden

OBJETIVO

- variables *inputs*
- predictoras
- variable independiente
- característica

Para encontrar

HACER

- variable *output*
- respuesta
- variable dependiente
- cualitativa o cuantitativa

PREDICCIONES

Regresión lineal con una variable

La distinción en el tipo de variable de salida ha llevado a una convención de nombre en las tareas de predicción:

- *regresión*: cuando se predicen *outputs* cuantitativas Y.
- *clasificación*: cuando se predicen *outputs* cualitativas G.

Estas dos tareas tienen mucho en común y, en particular, ambas se pueden ver como una tarea en aproximación de funciones.

Regresión lineal con una variable

Regresión lineal simple

La regresión lineal es un método matemático que modela la relación entre una variable dependiente, una variable independiente (regresores), y un término aleatorio ϵ.
La función de hipótesis tiene una forma general expresada por

$$h_\theta(x) = \theta_0 + \theta_1 x$$

El término θ_0 es el intercepto, también conocido como *bias* en machine learning.

Dado un vector de *inputs* $X^T = (X_1, X_2, \ldots, X_p)$, se predice el *ouput* Y vía el modelo

$$\hat{Y} = \hat{\theta}_0 + \sum_{j=1}^{p} X_j \theta_j$$

Usualmente conviene incluir la variable constante 1 en X, incluyendo $\hat{\theta}_0$ en el vector de coeficientes $\hat{\theta}$, escribiendo el modelo lineal como un producto interno

$$\hat{Y} = X^T \cdot \hat{\theta}$$

Regresión lineal
Modelo lineal múltiple

X^T denota la matriz transpuesta (en esta presentación los vectores se consideran vectores columna).

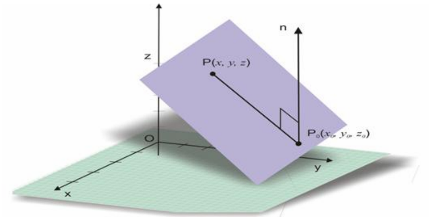
Aqui se ha modelado un simple $\hat{Y}$ *ouput*, o sea, $\hat{Y}$ es un escalar.

En genral $\hat{Y}$ puede ser un $K-$vector, en cuyo caso $\hat{\theta}$ sería una matriz de coeficientes de orden $p \times K$.

Regresión lineal

Modelo lineal múltiple, hiperplanos y conjuntos afines

En el espacio $(p+1)-$dimensional de *inputs*$-$*ouputs*, $(X, \hat{Y})$ representa un hiperplano.



Si la constante es incluida en X, entonces el hiperplano incluye el origen y es un subespacio.

Si no se incluye la constante es un conjunto afín que corta al eje Y en el punto $(0, \hat{\theta}_0)$.

Regresión lineal

Modelo lineal múltiple

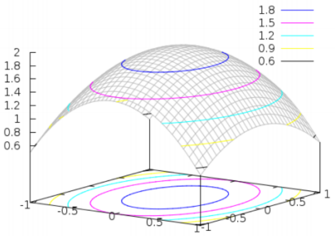
Visto como una función sobre el $p-$dimensional espacio de *input*,

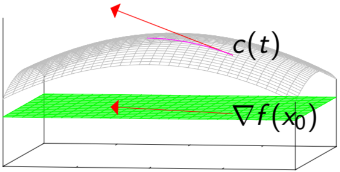
$$f(X) = X^T \theta$$

es lineal, y el gradiente

$$f'(X) = \theta$$

es un vector en el espacio de *inputs* que apunta en la dirección de ascenso más empinada.





Regresión lineal

Notación: de aquí en adelante se usará la siguiente convención

$$\mathbb{X} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} \dots x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} \dots x_n^{(2)} \\ & \vdots \\ x_1^{(i)} & x_2^{(i)} \dots x_n^{(i)} \\ & \vdots \\ x_1^{(m)} & x_2^{(m)} \dots x_n^{(m)} \end{bmatrix} = \begin{bmatrix} \left(x^{(1)}\right)^T \\ \left(x^{(2)}\right)^T \\ \vdots \\ \left(x^{(i)}\right)^T \\ \vdots \\ \left(x^{(m)}\right)^T \end{bmatrix} \qquad y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(i)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

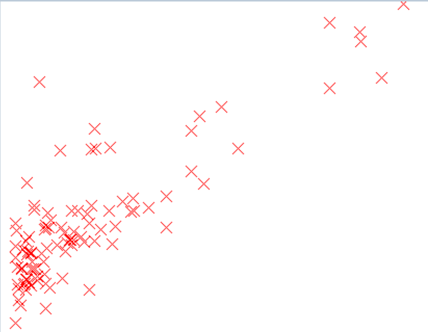
Cada dato es un vector en $\mathbb{R}^n$ el cual se representa por medio de cada fila de la matriz. n es el número de características de cada ejemplo y m es el número de ejemplos.

Regresión lineal con una variable

Ejemplo

Para comprender el proceso de regresión lineal, se realiza un ejemplo en una variable. $h_\theta = \theta_0 + \theta_1 x$

Ejemplo



Se desea encontrar una recta en el plano que mejor ajuste los datos

$$y = \theta_0 + \theta_1 x.$$

Definición

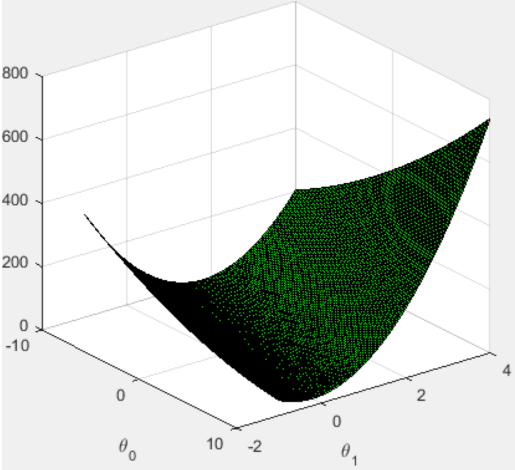
La función de costo sirve para medir la precisión de la función de hipótesis.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

J es llamada función de error cuadrático o error cuadrático medio. El promedio de los errores cuadráticos se divide por dos con el fin de ejecutar el algoritmo de gradiente descendiente de manera más funcional.

Regresión lineal con una variable

Superficie de la función de costo



Next Subsection

El algoritmo de gradiente descendiente se usa para encontrar el mínimo de la función de costos.
El algoritmo consiste en iterar hasta obtener convergencia

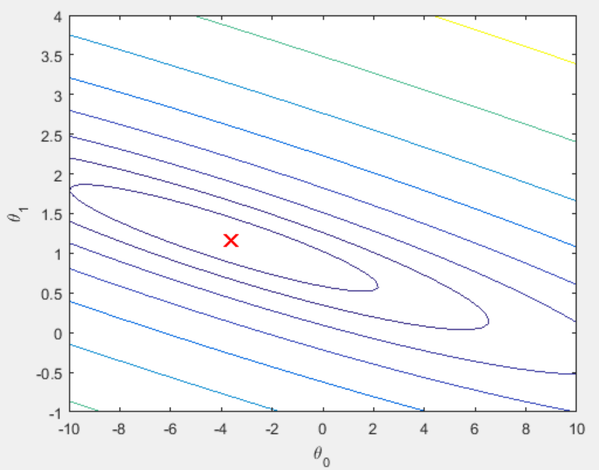
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

para $j = 0, 1$.

Regresión lineal con una variable

Gradiente descendiente

Contornos mostrando el mínimo de la función de costos.



ANFI
ANÁLISIS NUMÉRICO Y FINANCIERO

Next Subsection

Regresión lineal con una variable
Gradiente descendiente para regresión lineal

Repetir hasta obtener convergencia:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left((h_\theta(x^{(i)}) - y^{(i)})\, x^{(i)} \right)$$

Next Subsection

Regresión lineal con una variable

Ejemplos

Tomar los datos del archivo *datos_regresion_lineal_simple.csv* e implementar algoritmos para

- Hallar los parámetros de la regresión.
- Graficar los datos y la recta de regresión encontrada.
- Graficar la función de costos.
- Graficar los contornos de la superficie de la función de costos.

Observe que la función de costos se puede visualizar para el caso de regresión lineal simple, para dos o más variables independientes no es posible realizar tal visualización.

Next Subsection

Regresión lineal con múltiples características
Normalización

Cuando sea necesario normalizar los datos se debe

- Sustraer el valor promedio de cada característica del conjunto en el conjunto de datos.
- Después de sustraer la media, se escala (se divide) el valor de las características por su respectiva desviación estándar.

La desviación estándar es una manera de medir la variación presente en un rango de valores de una caraterística en particular.

Next Subsection

Regresión lineal con múltiples características
Gradiente descendiente

El algoritmo de gradiente descendiente implementado para el caso de una variable se puede generalizar al caso múltiple.
En el caso multivariado la función de costos se puede escribir de manera matricial como

$$J(\theta) = \frac{1}{2m}(X\,\theta - \vec{y}\,)^T(X\,\theta - \vec{y}\,)$$

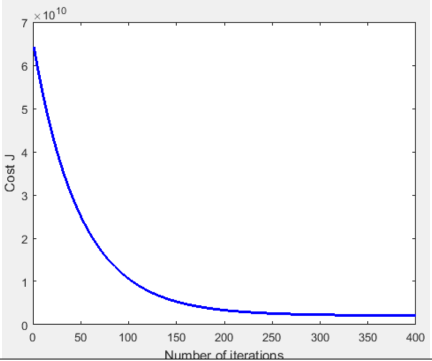
donde

$$X = [(x^{(i)})^T] \qquad \vec{y} = [y^{(i)}]$$

Regresión lineal con múltiples características

Selección de la tasa de aprendizaje α

Se deben probar diferentes tasas de aprendizaje en el conjunto de datos para encontrar una tasa de aprendizaje que converja rápidamente. Si α es muy grande $J(\theta)$ puede diverger. Si α es muy pequena el gradiente descendiente demora en llegar al óptimo.



La curva de convergencia cambia cuando se cambia la tasa de aprendizaje.

- Usando la mejor tasa de convergencia que usted encuentre. Ejecute el gradiente descendiente hasta converger y encuentre los valores finales de θ.
- Use los valores de θ para predecir los valores de la varible Y cuando son dados los valores de las características X.

Regresión lineal con múltiples características

Ejemplo regresión múltiple

Tomar los datos del archivo *datos_regresion_lineal_multiple_.csv* e implementar los algoritmos para

- Hallar los parámetros de la regresión.
- Graficar la curva de convergencia para diferentes tasas de aprendizaje α.

La curva de convergencia muestra el número de iteraciones contra los valores de $J(\theta)$. Para α en un rango apropiado la función de costos J decrece a medida que el número de iteraciones aumenta.

Next Subsection

Mínimos caudrados

Ajuste del modelo lineal

Para ajustar el modelo lineal a un conjunto de datos de entrenamiento por el método de mínimos cuadrados fijamos los coeficientes θ que minimizan la suma de cuadrados de los residuos

$$RSS(\theta) = \sum_{1=1}^{M} (y_i - x_i^T \ \beta)^2.$$

$RSS(\theta)$ es una funación cuadrática en los parámetros, y por tanto su mínimo siempre existe, pero puede que no sea único. En forma matricial

$$RSS(\theta) = (\vec{y} - X \ \beta)^T (\vec{y} - X \ \beta),$$

donde X es una matriz de orden $m \times n$ con cada fila un vector *inputs*, y $\vec{y}$ es un vector de *ouputs* de orden $m \times 1$ en el conjunto de entrenamiento.

Mínimos caudrados

Ecuaciones normales

Diferenciando con respecto a β en

$$RSS(\theta) = \sum_{1=1}^{M}(y_i - x_i^T\ \beta)^2,$$

se obtienen las ecuaciones normales:

$$X^T(\vec{y} - X \cdot \beta) = 0.$$

Si $X^T\ X$ es no singular, entonces la única solución es

$$\hat{\beta} = (X^T\ X)^{-1}\ X^T\ \vec{y}.$$

Mínimos caudrados
Ecuaciones normales

El valor que se predice para la $i-$ésima entrada x_i es

$$\hat{y}_i = \hat{y}(x_i) = x_i^T \cdot \hat{\beta}.$$

Y para un arbitrario *input* x_0, la predicción es

$$\hat{y}(x_0) = x_0^T \cdot \hat{\beta}.$$