Inteligencia Artificial

Programa de Ingeniería de Sistemas

Tema: Agentes que Aprenden - Regresión Lineal

LuzStellaRoblesPedrozo

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

Agentes que Aprenden

Los agentes inteligentes necesitan aprender del ambiente para poder adaptarse a él. Por lo tanto, la Inteligencia Artificial estudia el proceso de *Adquisición del Conocimiento* y el *Aprendizaje Automático*.

El *Aprendizaje Automático* es el área de la Inteligencia Artificial que se refiere al estudio, diseño y análisis de los sistemas que pueden aprender a partir de información.

En ML, existen dos tipos de aprendizaje:

- Supervisado
- No Supervisado

Agentes que Aprenden

Tipos de Aprendizaje

- Aprendizaje Supervisado

  Adquisición de conocimiento, por experiencia, datos del pasado y datos actuales, a través de pares de datos del tipo atributos-etiquetas.

- Aprendizaje no Supervisado

  Adquisición del conocimiento a través de información por atributos, que tiene como finalidad encontrar similitudes entre los datos.

- Aprendizaje por Refuerzo

  Adquisición del conocimiento basado en utilidades/castigos, que a través del tiempo generan un patrón de datos.

- Aprendizaje semi-supervisado

  Adquisición de conocimiento mediante técnicas híbridas entre aprendizaje supervisado y no supervisado.

Agentes que Aprenden

Espacio de Hipótesis

El problema de aprendizaje se puede re-escribir como un problema de búsqueda y optimización, en términos del espacio de hipótesis.

- Hipótesis

  Una Hipótesis se define como una posible solución al problema de aprendizaje. En general, para:

  — El aprendizaje supervisado, una hipótesis es un *Modelo*

  — El aprendizaje no supervisado, una hipótesis es un *Clasificador*

  — El aprendizaje por refuerzo, una hipótesis es una *Función de Costo*

- Espacio de Hipótesis

  Un espacio de hipótesis es el conjunto de hipótesis posibles para un problema de aprendizaje.

Aprendizaje Supervisado

La forma tradicional de Representación de Conocimiento de **Aprendizaje Supervisado** es mediante *Modelos*.

Algunas técnicas de Aprendizaje Supervisado:

- Regresión
- Redes Neuronales Artificiales
- Árboles de Decisión
- Clasificador Naïve Bayes
- Redes Bayesianas

Aprendizaje Supervisado

Aplicaciones Notables de **Aprendizaje Supervisado**:

- Modelos de Inferencia
- Modelos de Clasificación
- Filtrado de Datos
- Identificación de Sistemas
- Reconocimiento Facial
- Control Adaptable
- Modelado de Sistemas
- Predicción de Datos

Aprendizaje Supervisado

En Aprendizaje Supervisado se cuenta con:

- Un conjunto de datos –dataset–
- Y se sabe cuál es el resultado correcto o esperado

Por ejemplo, podría ser información histórica con variables atmosféricas:

- De un año
- Para una ciudad, y
- Si ha llovido o no en cada día

Aprendizaje Supervisado

En este ejercicio, los datos de entrada podría ser:

- Presión Atmosférica,
- Temperatura,
- Humedad Relativa,
- Velocidad, y
- Dirección del viento

La variable de salida sería:

- Con precipitación, o
- Sin precipitación

LuzStellaRoblesPedrozo

Aprendizaje Supervisado

El aprendizaje supervisado, se categoriza a su vez en:

- Regresión (Regression)
- Clasificación (Classification)

Aprendizaje Supervisado

En el **aprendizaje supervisado** de tipo ‘regresión’, se intenta *predecir* resultados dentro de una salida *continua*, es decir, un valor numérico.

En esta categoría podemos encontrar problemas de tipo ‘Predecir el precio de un bien raíz’, en ese caso, el precio será un *número*, por lo tanto es del tipo *regresión*.

Aprendizaje Supervisado

En el **aprendizaje supervisado** del tipo 'clasificación', a diferencia del tipo regresión, se intenta predecir resultados discretos.

Ejemplo: dado un conjunto de datos de entrada, se intenta predecir una categoría o etiqueta, así:

- Si, No.
- Comprar, Vender
- Vertebrado o Invertebrado

Regresión Lineal

Definición

La Regresión Lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes X_i y un término aleatorio ε (desde el punto de vista estadístico).

Regresión Lineal

Por qué usar Regresión Lineal?

- Se cuenta con unos datos
- Bajo el supuesto de que el conjunto de datos representan una curva
- Entender la curva en principio no es sencillo
- Sería sencillo si todo pudiera expresarse de la forma $y = mx + b$
- La anterior es una expresión lineal con dos variables y dos parámetros
- Se trabajan luego métodos para convertir un modelo no lineal a algo lineal
- A la aplicación de estos métodos es lo que se llama *Regresión*

Regresión Lineal

Por qué usar Regresión Lineal?

Continúa…

- $y = x^2 + 3$ este modelo no es lineal (es parabólico)
- Lo que se hace es linealizar la curva convirtiendo, por ejemplo, x^2 = f(z). De tal forma que ahora se debe graficar y versus f(z), y esto si genera línea recta
- Con la nueva función (linealizada) y el conjunto de datos, se intenta explicar el comportamiento de los mismos, pero a través de una línea recta
- Sin embargo, este método no puede ser aplicado siempre porque hay comportamientos que no son linealizables

Regresión Lineal

Desde Machine Learning

En los problemas de regresión, se reciben variables de entrada que luego genera una variable de salida y continua.

Debido a que la regresión es una categoría dentro del aprendizaje supervisado, se puede tener una idea del valor esperado.

Regresión Lineal

Desde Machine Learning

Para poder predecir, ya sea un valor discreto o continuo, se utiliza una función, llamada en Machine Learning ‘Hipótesis’

La función hipótesis, tiene la siguiente forma básica:

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

Y de forma general:

$$h(x) = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

Regresión Lineal

En Machine Learning

- Un algoritmo de regresión produce salidas que pertenecen a $\mathbb{R}$.
- A través de un proceso de entrenamiento, busca modelar lo mejor posible el comportamiento de un sistema real.
- La expresión

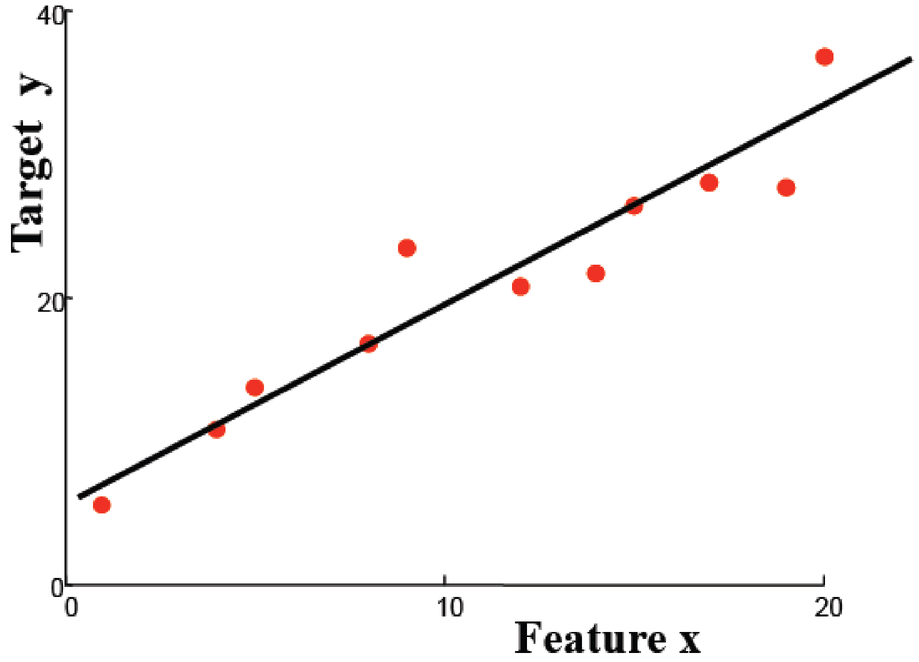
$$h(x) = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

- es lineal (una recta) en teta, de ahí el nombre de regresión lineal.

Donde θ_n representa cada uno de los parámetros del modelo (también llamados *pesos*).

Regresión Lineal

En Machine Learning



Regresión Lineal

En Machine Learning

En esta expresión:

$$h(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

x representan las entradas al sistema

θ_n representa cada uno de los parámetros del modelo (también llamados *pesos*)

Ejemplos de entradas al sistema: Edad, Salario, Número de años en el trabajo, etc.

Regresión Lineal

Aprendizaje

- Queremos obtener la recta que mejor capture el comportamiento de los datos entrenamiento.
- El primer paso para lograrlo es definir una *medida de error* que nos permita elegir cuál es la mejor recta de todas.
- En este caso queremos una recta que pase lo más cerca posible de todos los puntos de entrenamiento, por lo que hacemos la introducción de la siguiente *medida de error*:

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

Regresión Lineal

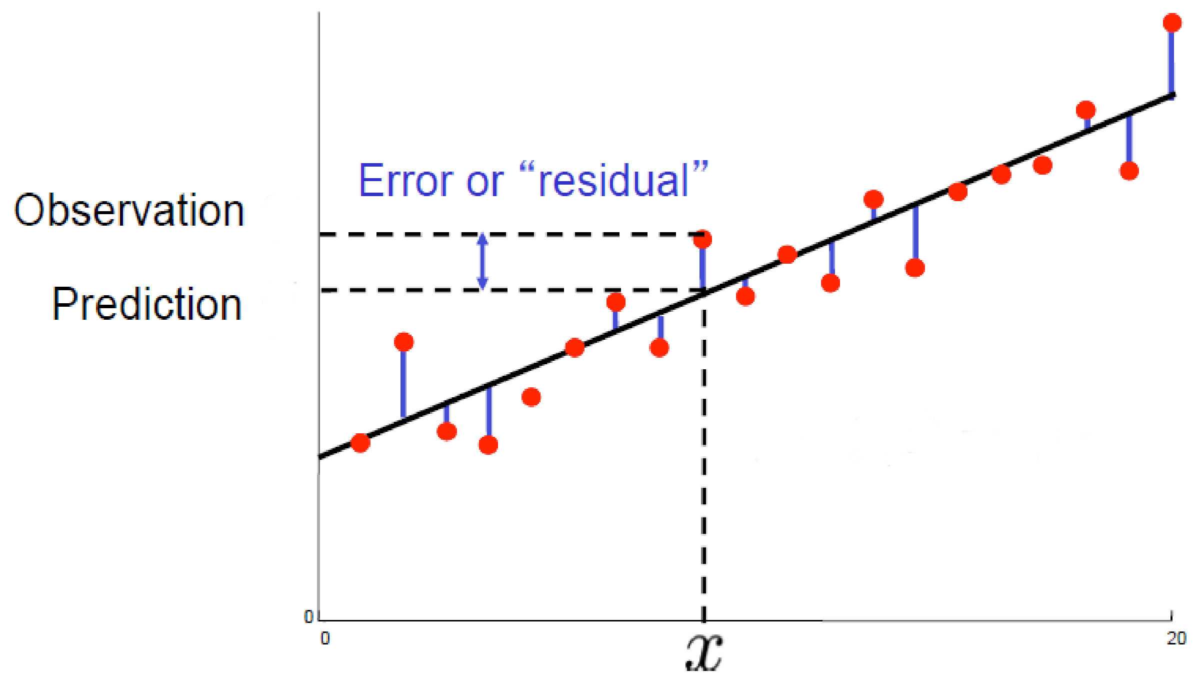
Aprendizaje

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

- Es de resaltar que la *medida de error* es una función de los parámetros ϴ.
- De aquí se puede concluir que el proceso de entrenamiento es realmente buscar un conjunto de ϴ que minimice el error *J(ϴ).*

Regresión Lineal

Aprendizaje



Regresión Lineal

Aprendizaje

Medida de error:

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

Donde:

m es la cantidad de elementos
(i) es un índice en el conjunto de entrenamiento
x es la matriz de entrada de variables
y es un vector de salida de resultado
variable que se entrena para predecir

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

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**, haciendo uso de Ecuaciones Normales

  En este método se realiza la minimización de forma explícita sin recurrir a un algoritmo interactivo.

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

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**, haciendo uso de Ecuaciones Normales

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

Si se observa detenidamente la estructura de la función de error, podremos notar que es una función cuadrática en θ.

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**, haciendo uso de Ecuaciones Normales

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

Recordemos que:

$$h(x) = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

El hecho de que sea cuadrática es de gran ayuda, porque nos garantiza la existencia de un único mínimo global!

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**

  Pero antes, es necesario agregar una nueva variable artificial $x_0 = 1$ (para incluir en la matriz el término intercept, y expresar la hipótesis de la siguiente manera:

$$h(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

$$h(x) = \sum_{i=0}^{n} \theta_i x_i$$

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**

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

En la anterior expresión, $h_\theta(x^{(i)}) = (x^{(i)})^T\theta = \theta^T \mathcal{X}^{(i)}$

Re-escribiendo:

$$min_\theta J(\theta) = \frac{1}{2}\sum_{n=1}^{N}\left(\theta . x^{(i)\,T} - y^{(i)}\right)^2$$

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**

  Dando un conjunto de entrenamiento, la matriz X pasa a ser la matriz de *m*n* (en realidad *m*n* + 1, si incluimos el término intercept) que contiene en sus filas, los valores de entrada para el entrenamiento.

  Re-escribiendo, usando la forma matricial:

$$\underline{\theta} = [\theta_0, \ldots, \theta_n]$$

$$\underline{y} = \left[y^{(1)} \ldots, y^{(m)}\right]$$

$$X = \begin{bmatrix} — (x^{(1)})^T — \\ — (x^{(2)})^T — \\ \vdots \\ — (x^{(m)})^T — \end{bmatrix}$$

Regresión Lineal

Aprendizaje

Para encontrar el conjunto de θ's que minimizan el error (la función de costo), se puede optar por:

- **El método Analítico**

  Adicionalmente, se tiene el vector $\vec{y}$ con *m* registros, que contiene todos los valores objetivo del conjunto de entrenamiento:

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

Regresión Lineal

Aprendizaje

- **El método Analítico**

  Es posible expresar y confirmar que $X\theta - \vec{y}$ es igual a:

$$
\begin{aligned}
X\theta - \vec{y} &= \begin{bmatrix} (x^{(1)})^T\theta \\ \vdots \\ (x^{(m)})^T\theta \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix} \\
&= \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \\ \vdots \\ h_\theta(x^{(m)}) - y^{(m)} \end{bmatrix}.
\end{aligned}
$$

Regresión Lineal

Aprendizaje

- **El método Analítico**

  Adicionalmente, utilizando el hecho de que para ***un*** vector z, se tiene que:

$$z^T z = \sum_i z_i^2$$

Regresión Lineal

Aprendizaje

El método Analítico

La medida de error puede ser re-escrita como:

$$min_{\theta}\, J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad = \quad \frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y})$$

Regresión Lineal

Aprendizaje

El método Analítico

La medida de error puede ser re-escrita como:

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

Aplicando la derivada:

$$min_{\theta} J(\theta) = \frac{1}{2}(X^T(X\theta - y) + (X\theta - y)^T X)$$

$$min_{\theta} J(\theta) = \frac{1}{2}\ (X^T X\theta - X^T y + \ \theta^T X^T X - y^T X\Big)$$

Regresión Lineal

Aprendizaje

El método Analítico

Aplicando la derivada:

$$min_{\theta} J(\theta) = \frac{1}{2}\ \left(X^T X\theta - X^T y + \ X^T X\theta^T - y^T X\right)$$

La traspuesta de un escalar es igual al escalar!

$$min_{\theta} J(\theta) = \frac{1}{2}\ \left(X^T X\theta - X^T y + \ X^T X\theta - y^T X\right)$$

El producto de $X^T y$ = $y^T X$

$$min_{\theta} J(\theta) = \frac{1}{2}\ \left(2X^T X\theta - 2X^T y\right)$$

Regresión Lineal

Aprendizaje

El método Analítico

Aplicando la derivada:

$$min_{\theta}J(\theta) = \frac{1}{2}\ \left(X^T X\theta - X^T y + \ X^T X\theta^T - y^T X\right)$$

La traspuesta de un escalar es igual al escalar!

$$min_{\theta}J(\theta) = \frac{1}{2}\ \left(X^T X\theta - X^T y + \ X^T X\theta^T - y^T X\right)$$

El producto de $X^T y$ = $y^T X$

$$min_{\theta}J(\theta) = \frac{1}{2}\ \left(2X^T X\theta - 2X^T y\right)$$

Regresión Lineal

Aprendizaje

El método Analítico

$$min_{\theta}J(\theta) = \frac{1}{2}\ \left(2X^{T}X\theta - 2X^{T}y\right)$$

$$min_{\theta}J(\theta) = \frac{1}{2}\ 2 * \left(X^{T}X\theta - X^{T}y\right)$$

$$min_{\theta}J(\theta) = (X^{T}X\theta - X^{T}y)$$

$$X^{T}X\theta\ =\ X^{T}y$$

$$\theta\ =\ (X^{T}X)^{-1}X^{T}y$$

Actividad Extra-clase

Taller de Regresión Lineal

El taller consiste en la implementación de un modelo basado en regresión lineal para predecir el precio de viviendas. Suponga que Ud. Está vendiendo su casa y quiere saber cuál sería un buen precio. Una forma de saberlo sería recolectar información acerca de ventas recientes de viviendas, y a partir de ella construir un modelo de predicción de precios.

El archivo ***ex1data2.txt*** contiene un conjunto de datos de precios de viviendas en Portland, Oregon (USA). La primera columna representa el área (en pies cuadrados), la segunda el número de habitaciones, y la tercera es el precio de la casa.

Utilice el 80% del conjunto de datos para realizar el entrenamiento. Use el 20% restante para probar la exactitud del modelo entrenado. Para esto, compare la salida del modelo de regresión lineal con el precio real de cada casa utilizada en la prueba. Ud. deberá reportar el Error Cuadrado Medio del modelo sobre los datos usados en la prueba.

Para desarrollar el taller, debe desarrollar el taller en **Python**, y adicionalmente, debe entregar el código fuente escrito.

Referencias

Material de apoyo de la semana registrados en SAVIO

Gracias!