Inteligencia Artificial

Programa de Ingeniería de Sistemas

Tema: Transformaciones No-Lineales

LuzStellaRoblesPedrozo

Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

Modelos lineales de Regresión y Clasificación

Características

Hasta ahora hemos visto un par de modelos de regresión y clasificación, donde las hipótesis hacen parte del conjunto del dominio de hipótesis lineales o hiperplanos.

Ventajas

Los modelos lineales necesitan poca información para ser entrenados (con respecto a otros modelos más complejos).

Desventajas

Los modelos lineales son limitados en su expresividad.

Entonces, **qué hacer** para tratar problemas de regresión y clasificación en los cuales una línea recta no sea suficiente?

Respuesta

Hacer uso de transformaciones no lineales!

Regresión Lineal con transformaciones no lineales

Características

- El algoritmo de Regresión Lineal que vimos, buscaba en el espacio de hipótesis lineales, el mejor hiperplano para ajustar un conjunto de puntos de entrenamiento.
- Recordemos la clase de hipótesis que usamos en la Regresión Lineal:

$$h_\theta(x) = \theta^t x = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n$$

- Un Error Cuadrado Medio (MSE) muy alto nos indicaría que la hipótesis lineal es insuficiente para describir los datos.

Qué sucedería si considerásemos hipótesis de regresión no lineales?

Respuesta

En vez de un hiperplano, tendríamos una hipersuperficie genérica (cuadrática, cúbica, grado-n) de regresión.

Ejemplo de hipótesis no lineal

Datos

Tenemos los siguientes datos de entrenamiento:

x	$f(x)$
0,86	2,49
0,09	0,83
-0,85	-0,25
0,87	3,10
-0,44	0,87
-0,43	0,02
-1,10	-0,12
0,40	1,81
-0,96	-0,83
0,17	0,43

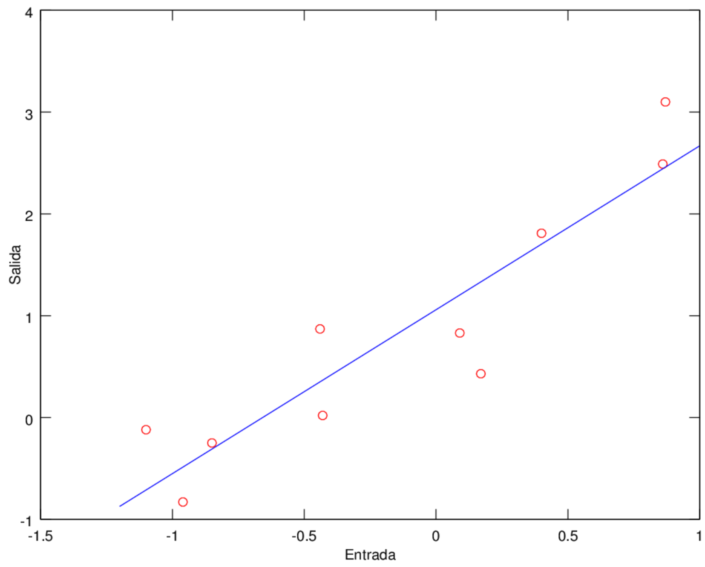
Ejemplo de hipótesis no lineal

Modelo y Gráfico

Ahora, aplicando el algoritmo de Regresión Lineal, obtenemos la siguiente hipótesis:

$$h(x) = 1,058 + 1,610x$$

$$\text{MSE} = 0,112$$



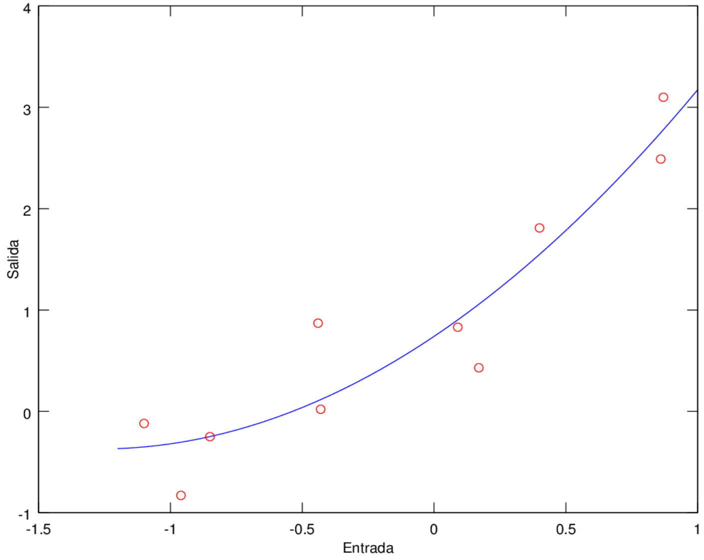
Ejemplo de hipótesis no lineal

Modelo y Gráfico

Probemos con una hipótesis cuadrática:

$$h(x) = 0,739 + 1,747x + 0,686x^2$$

$$\text{MSE} = 0,078$$



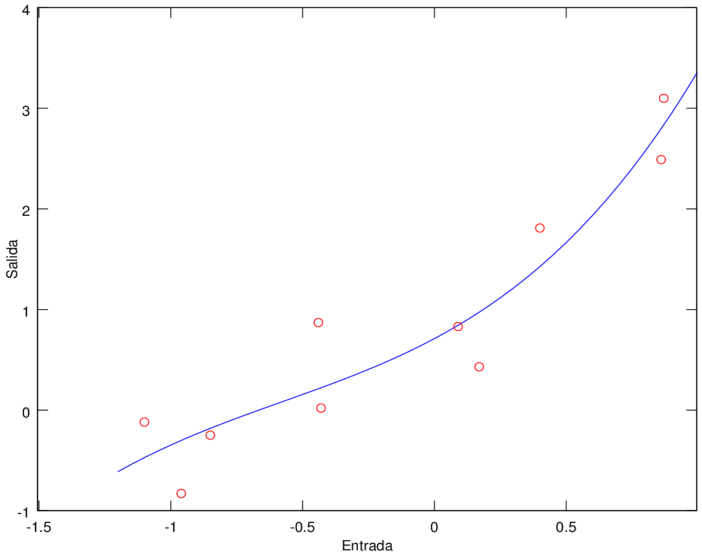
Ejemplo de hipótesis no lineal

Modelo y Gráfico

Por qué no una hipótesis cúbica?:

$$h(x) = 0,710 + 1,394x + 0,80x^2 + 0,464x^3$$

$$\mathrm{MSE} = 0,074$$

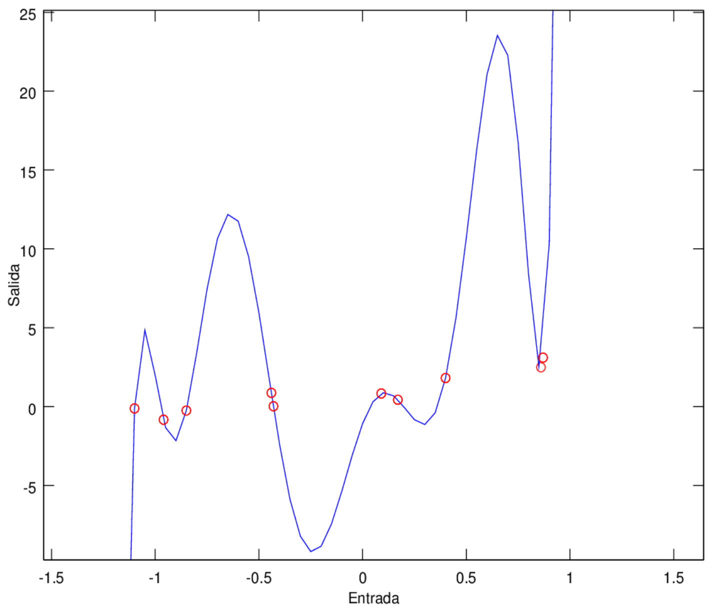


Ejemplo de hipótesis no lineal

Modelo y Gráfico

Por qué no una hipótesis de noveno grado?:

$$\text{MSE} = 3,7502 * 10^{-25}$$



Entrenamiento Regresión con features polinomiales

Características

- Vemos que entre más compleja sea la hipótesis, menor el MSE y más curva la gráfica.
- Sin embargo, no todo lo que brilla es oro. La disminución en el MSE puede ser a expensas de la habilidad de generalización del modelo.
- Cuál es la complejidad óptima? … más adelante, en el desarrollo del curso, se dará respuesta a esta pregunta.
- Por ahora concentrémonos en $\vec{\theta}$.

Entrenamiento Regresión con features polinomiales

Características

Cómo entrenamos nuestro modelos no lineales?

Respuesta

De la misma forma en que entrenamos nuestro modelo de Regresión Lineal.

Recordemos nuestra función de error (MSE):

$$J(\theta) = \frac{1}{N}\sum_{i=1}^{N}(h(x_i) - y_i)^2$$

Recordemos además que el entrenamiento cosiste en encontrar el vector θ que minimice $\mathcal{J}(\theta)$.

Entrenamiento Regresión con features polinomiales

Características

- Observemos que la hipótesis

$$h(x) = 0,739 + 1,747x + 0,686x^2$$

A pesar de ser cuadrática en x, sigue siendo LINEAL en θ!

- Y la función de error depende de θ, por lo que la derivación de la solución en un solo paso, lleva al mismo resultado:

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

- Por la misma razón, el algoritmo de Gradiente Descendente funciona sin ningún cambio, más allá del cálculo de h(x):

$$\theta j = \theta_j - \alpha \frac{1}{N} \sum_{i=1}^{N} (h(x_i) - y_i) x_{ij}$$

Entrenamiento Regresión con features polinomiales

Características

Para poder aplicar un modelo de Regresión polinomial, debemos completar nuestro conjunto de entrenamiento con los features del caso. Por ejemplo, si queremos usar el modelo cuadrático, nuestro conjunto de entrenamiento se convierte en:

x	x^2	$f(x)$
0,86	0,7396	2,49
0,09	0,0081	0,83
-0,85	0,7225	-0,25
0,87	0,7569	3,1
-0,44	0,1936	0,87
-0,43	0,1849	0,02
-1,1	1,21	-0,12
0,4	0,16	1,81
-0,96	0,9216	-0,83
0,17	0,0289	0,43

Entrenamiento Regresión con features polinomiales

Características

- Vemos entonces que no solo usamos los datos originales, sino transformaciones no lineales de las entradas, también llamadas features no lineales.
- El uso de dichas transformaciones incrementa sustancialmente la expresividad de nuestros modelos de regresión.
- De hecho, el uso de features no lineales es la clave de modelos de Machine Learning tales como: Redes Neuronales y las Máquina de Vectores de Soporte (SVMs).

CLASIFICACIÓN con features polinomiales

Características

- De igual manera, es posible utilizar clasificadores con fronteras de decisión no lineales.
- La buena noticia sigue siendo que el algoritmo de entrenamiento es prácticamente igual, con la diferencia de que $h_\theta(x)$ se calcula de forma diferente.

$$h_\theta(x) = \frac{1}{1+e^{-\theta^t x}}$$

$$\theta^T X = \theta_0 + \theta_1 X_1 + \theta_2 X_2$$

$$\theta^T X = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_1^2 + \theta_4 X_1^2 X_2 + \theta_5 X_2^2$$

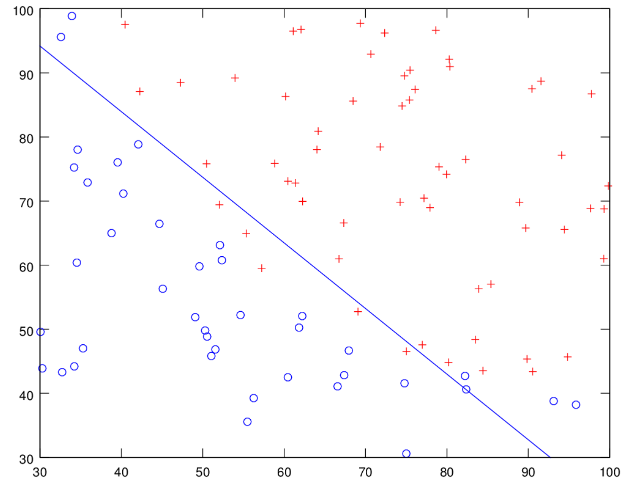
- La regla de actualización del gradiente descendente aplicado a la Regresión Logística con features polinomiales sigue siendo:

$$\theta j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(x_i) - y_i) x_{ij}$$

Entrenamiento Regresión Logística con features polinomiales

Características

- Recordemos nuestro ejemplo acerca de la predicción de la probabilidad de admisión de un estudiantes a la universidad.
- Habiendo entrenado el modelo, la frontera de decisión se ve de la siguiente manera:

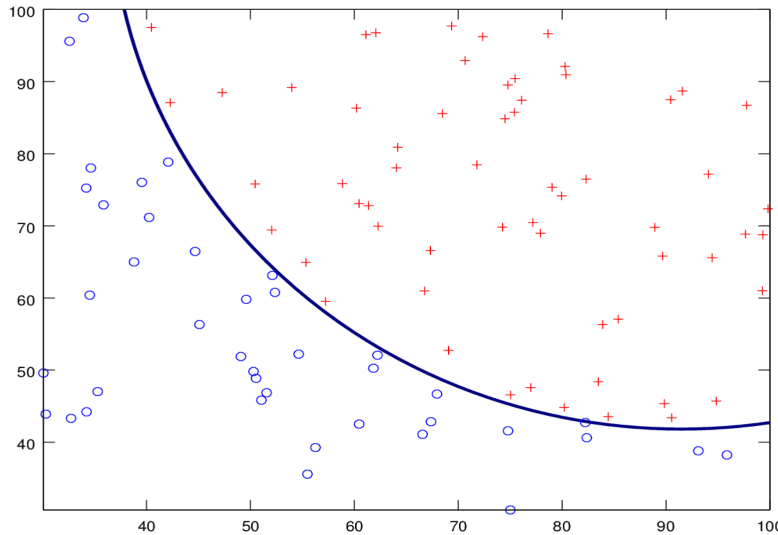


$$\theta^t x = -25,1612 + 0,2062 x_1 + 0,2014 x_2$$

Entrenamiento Regresión Logística con features polinomiales

Características

- Si utilizamos un modelo de segundo grado, este sería el resultado:

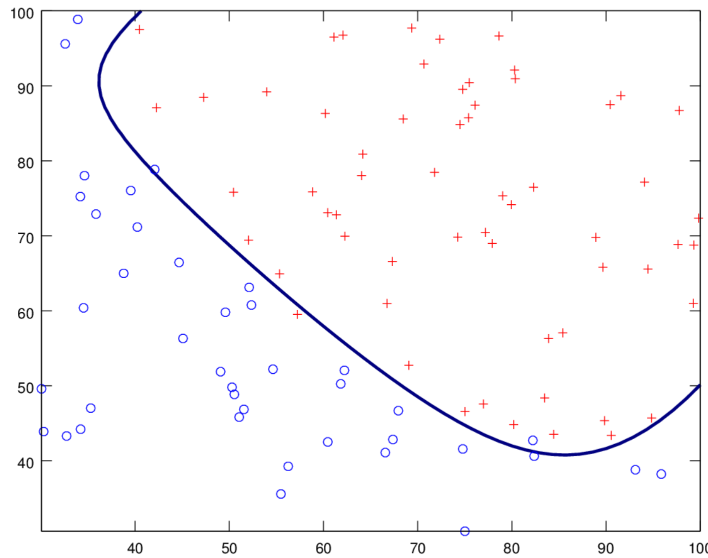


$$\theta^t x = -176,72 + 2,53x_1 + 1,768x_2 - 0,013x_1^2 - 0,007x_2^2$$

Entrenamiento Regresión Logística con features polinomiales

Características

- Si utilizamos un modelo de tercer grado, este sería el resultado:



$$\theta^t x = -0,206 - 3,037x_1 - 1,368x_2 + 0,071x_1^2 + 0,0043x_2^2 + 0,00042x_1^3 + 0,00026x_3^3$$

Entrenamiento Regresión Logística con features polinomiales

Características

- De igual manera que en la Regresión Lineal, entre más compleja sea la hipótesis de clasificación a utilizar, más complicada puede ser la frontera de decisión del clasificador.
- En este caso, un modelo cuadrático lleva el error de clasificación prácticamente a cero.
- La mala noticia es que la complejidad adicional viene con un costo de generalización.
- Existen resultados probabilísticos que señalan que: a más complejidad del clasificador, mayor error de generalización puede ser esperado.

Actividad Extra-clase

Taller 3

Enunciado en SAVIO!

Referencias

Material de apoyo de la semana registrados en SAVIO.

Gracias!