Introducción a Machine Learning

Sesión 2.1: Regresión Logistica y Regularización

Ronald Cárdenas Acosta

Agosto, 2016

Outline

1. Regresión Logística (Logistic Regression)
   - Definición
   - Función costo
   - Optimización
2. Regularización
   - Sub-ajuste (Under-fitting) y sobre-ajuste (Over-fitting)
   - Opciones de solución
   - Regularización: Definición
   - Modelos regularizados

Regresión Logística

- Usado para clasificación, a pesar del nombre "regresión"
- Solo considera clasificación binaria: $y \in 0, 1$
  - Para clasificación multi-clase se usa el enfoque One.vs.All
  - Regresor Logístico multiclase $\equiv$ Clasificadores de Máxima Entropía (usados ampliamente en NLP)
- Hipótesis de modelo $h_w(x)$:

$$h_w(x) = sigmoid(w.x) = \frac{1}{1+\exp(-w.x)}$$
$$0 \leq h_w(x) \leq 1$$

Outline

Función costo / función objetivo

$$L(w) = \frac{1}{N}\sum_{i=1}^{N} Cost(h_w(x^i), y^i) \tag{1}$$

Donde:

$$Cost(h_w(x), y) = \begin{cases} -log(h_w(x)) & \text{si } y = 1 \\ -log(1 - h_w(x)) & \text{si } y = 0 \end{cases}$$

$$Cost(h_w(x), y) = -y.log(h_w(x)) - (1-y).log(1 - h_w(x))$$

Función costo / función objetivo

- Reemplazando $Cost(h_w(x), y)$

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N}[y.log(h_w(x^i)) + (1 - y^i).log(1 - h_w(x^i))] \quad (2)$$

- Modelo lineal en el espacio logarítmico, es decir, log-lineal

Outline

1. Regresión Logística (Logistic Regression)
   - Definición
   - Función costo
   - Optimización
2. Regularización
   - Sub-ajuste (Under-fitting) y sobre-ajuste (Over-fitting)
   - Opciones de solución
   - Regularización: Definición
   - Modelos regularizados

Optimización: Gradient Descent

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N}[y.log(h_w(x^i)) + (1-y^i).log(1-h_w(x^i))]$$

- Objetivo: $\hat{w} = argminL(w)$
- Iterar para cada característica $j \in [1, M]$:

$$w_j = w_j - \alpha.\nabla_w L(w)$$
$$w_j = w_j - \alpha\frac{1}{N}\sum_{i=1}^{N}(h_w(x^i) - y^i)x_j^i$$

Optimización

- Otros algoritmos de optimización
  - Conjugate Gradient
  - BFGS
  - L-BFGS
  - ADAM, ADAGRAD (especiales para redes neuronales)
- Ventajas
  - No se necesita escoger α manualmente
  - Generalmente convergen más rápido que Gradient Descent
- Desventaja de ser más complejos, pero existen implementaciones disponibles en C++, Python, Java.

Outline

Sub-ajuste (Under-fitting) y sobre-ajuste (Over-fitting)

Sub-ajuste: Under-fitting

Cuando el modelo presenta alto "bias", es decir, los parametros reales y los parametors estimados difieren bastante.

$$Bias(\hat{w}) = E[\hat{w}] - w$$

Sobre-ajuste: Over-fitting

Cuando el modelo presenta alta varianza, es decir, un pequeño cambio a la data observada provoca grandes cambios en los parámetros estimados.

$$Var(\hat{w}) = E[(\hat{w} - E(\hat{w}))^2]$$

Ejemplo: Predicción de costo de una casa

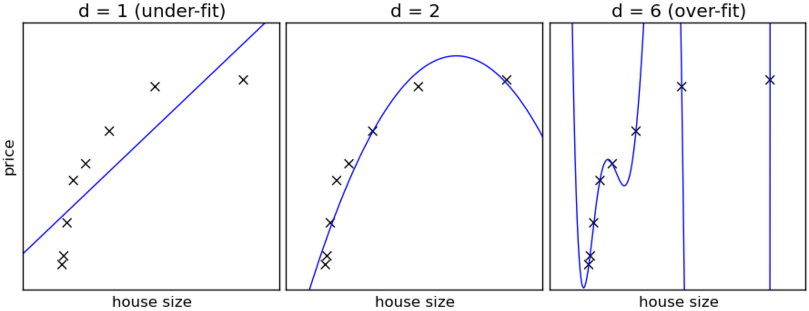

Figure: Modelo: ajuste de polinomio de grado d

¿Porqué sucede esto?

- Sub-ajuste
  - El modelo usa muy pocas *características* x_j
  - La data de entrenamiento es insuficiente.
- Sobre-ajuste
  - Usar demasiadas *características* x_j hace que el modelo "replique la data". No podrá generalizar bien para data nueva.

Outline

Opciones de solución

- Reducir o aumentar número de características según sea el caso.
- Regularización
  - Mantener el número de características, pero reducir la magnitud de los parámetros w_j
  - Funciona mejor cuando se tiene gran número de características (data X dispersa)

Outline

Regularización

- No afecta la definición de la hipótesis $h_w(x)$
- Se aplica a la función de costo:

$$L(w) = \frac{1}{N}\sum_{i=1}^{N} cost(h_w(x), y) + \lambda.R(w)$$

Donde:

- $R(w)$: término de regularización, definido por una *norma* matemática
- λ: parámetro de regularizacón. Controla el grado de regularización

Tipos de regularización

- Regularización L1: $R(w) = \|w\|_1 = \sum_{j=1}^{M} w_j$
- Regularización L2: $R(w) = \frac{1}{2}\|w\|_2^2 = \frac{1}{2} w^T w = \frac{1}{2} \sum_{j=1}^{M} w_j^2$
- Promediado de parámetros por época (usado en perceptrón y variantes)
- Dropout (usado para Deep Learning)
- Generalizaciones de normas para n dimensiones (p.e. norma Frobenius, norma nuclear)

Outline

1. Regresión Logística (Logistic Regression)
   - Definición
   - Función costo
   - Optimización

2. Regularización
   - Sub-ajuste (Under-fitting) y sobre-ajuste (Over-fitting)
   - Opciones de solución
   - Regularización: Definición
   - Modelos regularizados

Perceptron + Regularización

- Función de costo

$$L(x, y; w) = \sum_{i=1}^{M} 1 - [[y_{pred} == y]] + \frac{1}{2}\lambda\|w\|_2^2$$

- Optimización

$$w_j = w_j + \alpha x + \lambda w_j$$

Regresión Logística + Regularización

- Función de costo

$$L(w) = -\frac{1}{N}\sum_{i=1}^{N}[y.log(h_w(x^i)) + (1-y^i).log(1-h_w(x^i))] + \frac{1}{2N}\lambda\|w\|_2^2$$

- Optimización

$$w_j = w_j - \alpha[\frac{1}{N}\sum_{i=1}^{N}(h_w(x^i) - y^i)x_j^i - \frac{\lambda}{N}w_j]$$