Redes Neuronales Monocapa III

Modelos de la computación (Aprendizaje supervisado)

Francisco Fernández Navarro

Departamento de Lenguajes y Ciencias de la Computación
Área: Ciencias de la Computación e Inteligencia Artificial

Índice de contenidos

1. ADALINA
   - Introducción
   - Modelo funcional
   - Problema de optimización
   - Estimación de parámetros para la implementación en línea
   - Estimación de parámetros para la implementación por lotes
   - Ejemplos de implementación
   - Versión analítica de ADALINA
2. Neuronas con salida continua
   - Modelo funciónal
   - Problema de optimización y estimación de parámetros

ADALINA

Breve reseña histórica

- El término "“Adaline” es una abreviatura de ”Adaptive Linear Neuron” (Neurona Lineal Adaptativa).
- Modelo de red neuronal artificial que fue desarrollado por Bernard Widrow y Ted Hoff en Stanford University en la década de 1960.

Fundamentación

- El Adaline se desarrolló como una alternativa del Perceptrón de Frank Rosenblatt.
- A diferencia del Perceptrón, el Adaline utiliza una función de activación lineal y realiza ajustes continuos en los pesos de las conexiones entre las neuronas para minimizar el error de predicción.

Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

Modelo funcional

La salida de la unidad para un patrón n-ésimo, $\mathbf{x}_n \in \mathbb{R}^K$ es:

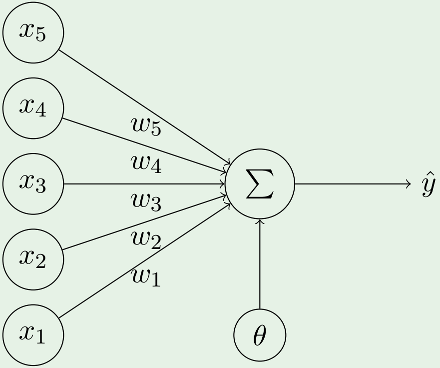
$$\hat{y}(\mathbf{x}_n) = f(\mathbf{x}_n) = \sum_{k=1}^{K} w_k x_{nk} - \theta, \tag{1}$$

donde $\mathbf{w} \in \mathbb{R}^K$ son los pesos con los que se ponderan los valores de entrada (conocidos como **pesos sinápticos**) y el parámetro θ es el **umbral o sesgo** de la unidad de procesamiento. Alternativamente:

$$\hat{y}(\mathbf{x}_n) = f(\mathbf{x}_n) = \sum_{k=1}^{K+1} w_k x_{nk}, \tag{2}$$

considerando una entrada adicional con valor $x_{nK+1} = -1$ $(\forall n, n = 1, \ldots, N)$ cuyo peso sináptico $w_{K+1} = \theta$.

ADALINA con 5 entradas



Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

ADALINA

Problema de optimización

- Los parámetros del modelo se estiman a través del siguiente problema de optimización:

$$\min_{\mathbf{w}\in\mathbb{R}^{K+1}} \quad E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 = \frac{1}{2}\sum_{n=1}^{N}\left(y_n - \sum_{k=1}^{K+1} w_k x_{nk}\right)^2$$

  donde $\mathbf{w} = (w_1, \ldots, w_{K+1})$ es el vector de pesos a estimar.
- Minimización error cuadrático medio.
- Se trata de determinar los pesos sinápticos que consiguen que las salidas de la red sean lo más parecidas a las salidas deseadas para el conjunto dado de patrones de entrenamiento.

Cuestiones

- ¿Porqué los errores de aproximación están al cuadrado?
- ¿Por qué se elevan al cuadrado los errores individuales al calcular el ECM en lugar de tomar el valor absoluto de los errores?
- ¿Cuál es el efecto de los valores atípicos (outliers) en el ECM? ¿Por qué el ECM puede ser sensible a los valores atípicos?
- ¿Cuáles son algunas ventajas y desventajas de utilizar el ECM como métrica de evaluación de un modelo de regresión?

Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

Estimación de parámetros. Cuestiones previas

- Vamos a seguir para la estimación un procedimiento iterativo, de tal forma que los parámetros serán modificados en unas iteraciones que irán de la 1 a la I, y en cada iteración los parametros serán modificados patrón a patrón, $n = 1, \ldots, N$.
- Para ello es necesario modificar la formulación de base para indicar que $\hat{y}_n(i)$, es la estimación del patrón n-ésimo en la iteración i-ésima, y que $w_k(i, n)$ es el valor del vector de pesos en la dimensión k-ésima, en la iteración i-ésima y en el patrón n-ésimo.

Estimación de parámetros. Cuestiones previas

- El patrón en la iteración i, en su dimensión k, será modificado debido al patrón n en función de la siguiente ecuación:

$$w_k(i, n+1) = w_k(i, n) + \Delta w_k(i, n), \tag{3}$$

- La varianción de dicho término quedará definida como:

$$\Delta w_k(i, n) = -\eta \frac{\partial E(\mathbf{w})}{\partial w_k(i, n)} = \eta (y_n - \hat{y}_n(i)) x_{nk},$$

donde $\hat{y}_n(i) = \sum_{k=1}^{K+1} w_k(i, n) x_{nk}$

- El parámetro η (**tasa de aprendizaje**) controla la longitud del paso que vamos a dar en la dirección opuesta del gradiente.

Flujo algorítmico - Algoritmo en línea (patrón a patrón)

ADALINA-EN LINEA $(\mathcal{D}, \eta > 0)$:

$\mathbf{X} \leftarrow (\mathbf{X}\ \ -\mathbf{1}_N) \in \mathbb{R}^{N\times(K+1)}$
for $k = 1$ until $K + 1$ **do**
    $w_k(1,0) \leftarrow 2 \times \text{rand}() - 1, 0 \leq \text{rand}() < 1$
    $(\mathbf{w}(1,0) = (w_1(1,0), \ldots, w_{K+1}(1,0)))$.
end for
for $i = 1$ until I **do**
    for $n = 1$ until N **do**
        $\hat{y}_n(i) \leftarrow \sum_{k=1}^{K+1} w_k(i,n) x_{nk}$
        for $k = 1$ until $K + 1$ **do**
            $\Delta w_k(i,n) \leftarrow \eta(y_n - \hat{y}_n(i)) x_{nk}$
            $w_k(i,n) \leftarrow w_k(i,n-1) + \Delta w_k(i,n)$.
        end for
    end for
end for

Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

Adalina por lotes

- **Idea**: Introducir los N patrones directamente y comparar las salidas de la red con las salidas deseadas, pasando entonces a actualizar los pesos sinápticos.
- La modificación de los pesos sinápticos se hace tomando como función de error el error medio, por lo que el problema de optimización queda como:

$$\min_{\mathbf{w}\in\mathbb{R}^{K+1}} \quad E(\mathbf{w}) = \frac{1}{2N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 = \frac{1}{2N}\sum_{n=1}^{N}\left(y_n - \sum_{k=1}^{K+1} w_k x_{nk}\right)^2$$

ADALINA

Adalina por lotes

- No necesitamos iterar por patrón por lo que tendrémos parámetros denotados como: $w_k(i)$ (valor del peso sináptico en su dimensión k en la iteración i).
- La regla de aprendizaje quedaría como:

$$\Delta w_k(i) = -\eta \frac{\partial E(\mathbf{w})}{\partial w_k(i)} = \frac{\eta}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n(i)) x_{nk},$$

Flujo algorítmico - Algoritmo por lotes

ADALINA-POR LOTES $(\mathcal{D},\, \eta > 0)$:

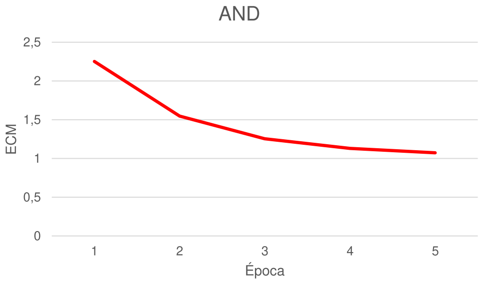
$\mathbf{X} \leftarrow (\mathbf{X}\ \ -\mathbf{1}_N) \in \mathbb{R}^{N\times(K+1)}$
for $k = 1$ until $K+1$ **do**
  $w_k(0) \leftarrow 2 \times \text{rand}() - 1$, $(\mathbf{w}(0) = (w_1(0), \ldots, w_{K+1}(0)))$.
end for
for $i = 1$ until I **do**
  for $k = 1$ until $K+1$ **do**
    $\Delta w_k(i) \leftarrow 0$
  end for
  for $n = 1$ until N **do**
    $\hat{y}_n(i) \leftarrow \sum_{k=1}^{K+1} w_k(i,n) x_{nk}$
    for $k = 1$ until $K+1$ **do**
      $\Delta w_k(i) \leftarrow \Delta w_k(i) + (y_n - \hat{y}_n(i)) x_{nk}$
    end for
  end for
  $\Delta w_k(i) \leftarrow \frac{\eta \Delta w_k(i)}{N}$
  for $k = 1$ until $K+1$ **do**
    $w_k(i) \leftarrow w_k(i-1) + \Delta w_k(i)$.
  end for
end for

Índice de contenidos

1. ADALINA
   - Introducción
   - Modelo funcional
   - Problema de optimización
   - Estimación de parámetros para la implementación en línea
   - Estimación de parámetros para la implementación por lotes
   - Ejemplos de implementación
   - Versión analítica de ADALINA
2. Neuronas con salida continua
   - Modelo funciónal
   - Problema de optimización y estimación de parámetros

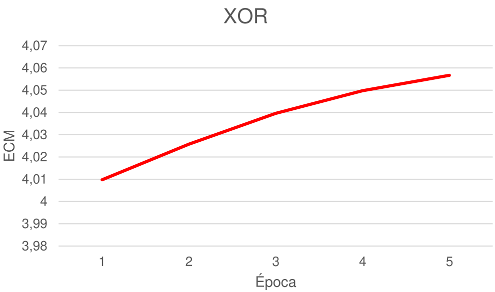
Ejemplo ADALINA AND

t	x1	x2	-1	Z	Y	w1	w2	rho
						0	0	0
1	1	1	-1	1	0,00	0,10	0,10	-0,10
2	1	-1	-1	-1	0,10	-0,01	0,21	0,01
3	-1	1	-1	-1	0,21	0,11	0,09	0,13
4	-1	-1	-1	-1	-0,33	0,18	0,16	0,20
5	1	1	-1	1	0,14	0,26	0,24	0,11
6	1	-1	-1	-1	-0,09	0,17	0,33	0,20
7	-1	1	-1	-1	-0,04	0,27	0,24	0,30
8	-1	-1	-1	-1	-0,80	0,29	0,26	0,32
9	1	1	-1	1	0,23	0,37	0,33	0,24
10	1	-1	-1	-1	-0,21	0,29	0,41	0,32
11	-1	1	-1	-1	-0,19	0,37	0,33	0,40
12	-1	-1	-1	-1	-1,10	0,36	0,32	0,39
13	1	1	-1	1	0,29	0,43	0,39	0,32
14	1	-1	-1	-1	-0,28	0,36	0,47	0,39
15	-1	1	-1	-1	-0,28	0,43	0,39	0,46
16	-1	-1	-1	-1	-1,28	0,40	0,37	0,43
17	1	1	-1	1	0,33	0,47	0,43	0,37
18	1	-1	-1	-1	-0,33	0,40	0,50	0,43
19	-1	1	-1	-1	-0,34	0,47	0,43	0,50
20	-1	-1	-1	-1	-1,40	0,43	0,39	0,46



Ejemplo ADALINA XOR

t	x1	x2	-1	Z	Y	w1	w2	rho
						0	0	0
1	1	1	-1	-1	0,00	-0,10	-0,10	0,10
2	1	-1	-1	1	-0,10	0,01	-0,21	-0,01
3	-1	1	-1	1	-0,21	-0,11	-0,09	-0,13
4	-1	-1	-1	-1	0,33	0,02	0,04	0,00
5	1	1	-1	-1	0,06	-0,08	-0,06	0,11
6	1	-1	-1	1	-0,13	0,03	-0,18	0,00
7	-1	1	-1	1	-0,20	-0,09	-0,06	-0,12
8	-1	-1	-1	-1	0,27	0,04	0,07	0,00
9	1	1	-1	-1	0,10	-0,07	-0,04	0,11
10	1	-1	-1	1	-0,15	0,04	-0,15	0,00
11	-1	1	-1	1	-0,19	-0,08	-0,03	-0,12
12	-1	-1	-1	-1	0,23	0,04	0,09	0,00
13	1	1	-1	-1	0,13	-0,07	-0,02	0,12
14	1	-1	-1	1	-0,16	0,05	-0,14	0,00
15	-1	1	-1	1	-0,19	-0,07	-0,02	-0,12
16	-1	-1	-1	-1	0,21	0,05	0,10	0,00
17	1	1	-1	-1	0,15	-0,06	-0,01	0,12
18	1	-1	-1	1	-0,17	0,05	-0,13	0,00
19	-1	1	-1	1	-0,18	-0,07	-0,01	-0,12
20	-1	-1	-1	-1	0,20	0,05	0,11	0,00



Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

Adalina versión no iterativa

- Vamos a estimar los parámetros sin la necesidad de iterar.
- El conjunto de entrada será extendido con un vector columna de -1s al final de la matriz original ($\mathbf{X} = (\mathbf{X} \ \ -\mathbf{1}_N) \in \mathbb{R}^{N \times (K+1)}$).
- La variable dependiente en forma matricial será definida como $\mathbf{Y} = (y_1, \ldots, y_n) \in \mathbb{R}^N$.
- El problema de optimización quedaría como:

$$\min_{\mathbf{w} \in \mathbb{R}^{K+1}} \quad E(\mathbf{w}) = ||\mathbf{Y} - \mathbf{X}\mathbf{w}||^2 = (\mathbf{Y} - \mathbf{X}\mathbf{w})'(\mathbf{Y} - \mathbf{X}\mathbf{w})$$

donde $\mathbf{w} = (w_1, \ldots, w_K, \theta) \in \mathbb{R}^{K+1}$.

ADALINA

Adalina versión no iterativa

- Expandimos la función de error:

$$||\mathbf{Y} - \mathbf{X}\mathbf{w}||^2 = \mathbf{Y}'\mathbf{Y} - 2\mathbf{w}'\mathbf{X}'\mathbf{Y} + \mathbf{w}'\mathbf{X}'\mathbf{X}\mathbf{w}$$

- Derivamos e igualamos a cero:

$$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{w} = 0,$$

  por lo que

$$\mathbf{X}'\mathbf{X}\mathbf{w} = \mathbf{X}'\mathbf{Y},$$

  y por tanto

$$\mathbf{w} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Flujo algorítmico - Algoritmo versión analítica

ADALINA-ANALITICO $(\mathcal{D})$:

1: $\mathbf{X} \leftarrow (\mathbf{X} \;\; -\mathbf{1}_N) \in \mathbb{R}^{N\times(K+1)}$

2: $\mathbf{w} \leftarrow (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

Índice de contenidos

1. ADALINA
   - Introducción
   - Modelo funcional
   - Problema de optimización
   - Estimación de parámetros para la implementación en línea
   - Estimación de parámetros para la implementación por lotes
   - Ejemplos de implementación
   - Versión analítica de ADALINA
2. Neuronas con salida continua
   - Modelo funciónal
   - Problema de optimización y estimación de parámetros

Neuronas de salida continua

Modelo funcional

- La salida de la unidad para un patrón n-ésimo, $\mathbf{x}_n \in \mathbb{R}^K$ es:

$$\hat{y}(\mathbf{x}_n) = g\left(\sum_{k=1}^{K+1} w_k x_{nk}\right), \tag{4}$$

  donde $\mathbf{w} \in \mathbb{R}^K$ son los pesos con los que se ponderan los valores de entrada y $g(.)$ es la función de transferencia.
- La función de transferencia va a ser una función diferenciable y no decreciente.

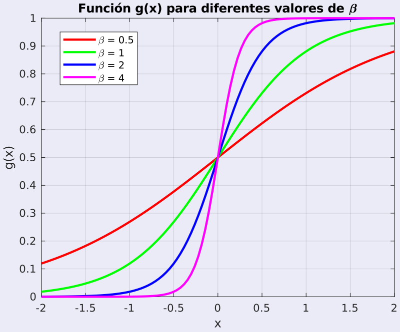
Neuronas de salida continua

Tipos de función de transferencia

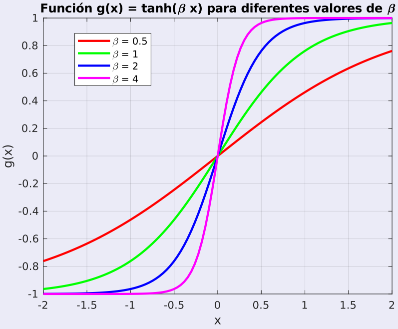
- La función logística: $g(x) = \frac{1}{1+\exp(-2\beta x)}$. Función de aplastamiento, $g : \mathbb{R} \to [0,1]$. El parámetro de ganancia β controla la pendiente de la curva. Derivada muy simple: $g'(x) = 2\beta g(x)(1 - g(x))$.
- La función tangente hiperbólica: $g(x) = \tanh(\beta x) = \frac{e^{\beta x} - e^{-\beta x}}{e^{\beta x} + e^{-\beta x}}$. Función de aplastamiento, $g : \mathbb{R} \to [-1,1]$. Derivada muy simple: $g'(x) = \beta(1 - g(x)^2)$

Representación gráfica - Funciones Logística y Tangente Hiperbólica

Función Logística



Función Tangente Hiperbólica



Índice de contenidos

1 ADALINA
- Introducción
- Modelo funcional
- Problema de optimización
- Estimación de parámetros para la implementación en línea
- Estimación de parámetros para la implementación por lotes
- Ejemplos de implementación
- Versión analítica de ADALINA

2 Neuronas con salida continua
- Modelo funciónal
- Problema de optimización y estimación de parámetros

Nodos de salida continua

Problema de optimización

- Los parámetros del modelo se estiman a través del siguiente problema de optimización:

$$\min_{\mathbf{w}\in\mathbb{R}^{K+1}} \quad E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2 = \frac{1}{2}\sum_{n=1}^{N}\left(y_n - g\left(\sum_{k=1}^{K+1} w_k x_{nk}\right)\right)^2.$$

- Por lo que en el aprendizaje en línea, la variación viene definida como:

$$\Delta w_k(i,n) = -\eta\frac{\partial E(\mathbf{w})}{\partial w_k(i,n)} = \eta(y_n - \hat{y}_n(i))g'\left(\sum_{k=1}^{K+1} w_k x_{nk}\right)x_{nk}.$$

- Para el aprendizaje por lotes usamos de función de error el error cuadrático promedio (al igual que hicimos con el ADALINA).

¡Gracias por vuestra atención!