A continuación, se realiza la derivación del gradiente para un perceptrón multicapa.

SUPUESTOS USADOS EN LA DERIVACIÓN DEL GRADIENTE DE LA FUNCIÓN DE ERROR:

1. Se limita un máximo de 3 capas de procesamiento (2 capas ocultas)
2. Ecuaciones para perceptrones más simples (menos capas) se obtienen directamente de la formulación.
3. Los contadores para las capas de procesamiento 0, 1, 2 y 3 son y respectivamente.
4. es la salida de una neurona de la capa . Si es la última capa, entonces equivale a la salida del modelo. es la entrada al perceptrón ().
5. es la entrada a una neurona de la capa .
6. son los pesos de la capa hacia la capa .
7. son los pesos de la neurona bias hacia cada una de las neuronas de procesamiento de la capa
8. es la activación de la neurona y es su derivada para una neurona de la capa .

ECUACIONES BÁSICAS:

; ; ;

; ;

DERIVACIÓN ANALÍTICA DEL GRADIENTE:

es el criterio de error utilizado. La derivación es independiente de la función específica utilizada. Por simplicidad se usará la sumatoria del error cuadrático para el patrón actual:

Entonces:

La sumatoria anterior sólo es diferente de cero para .

Gradiente para la 3ª capa de procesamiento.

Gradiente para la 2ª capa de procesamiento (note la estructura recursiva de las ecuaciones).

Gradiente para la 1ª capa de procesamiento

Para un perceptrón de una sola capa de procesamiento (monocapa) con entradas y neuronas de salida:

Regla de aprendizaje: