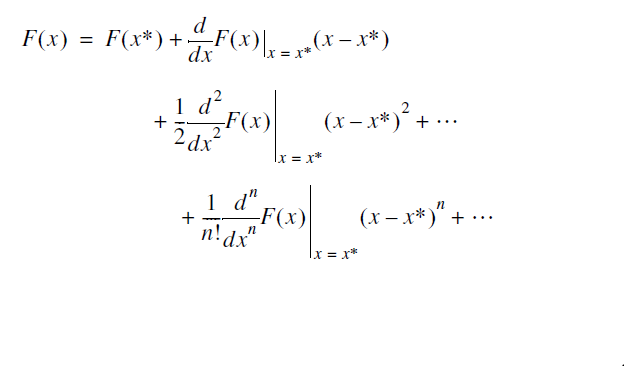
Resumen cap 8

En este capítulo se analiza a profundidad algo llamado “Aprendizaje por rendimiento” (Perfomance Learning) el cual destaca por cambiar los valores de pesos y bias para poder encontrar los valores máximos y mínimos para tener una mejor optimización de los resultados obtenidos.

Primero que nada tenemos que saber que existen 2 cosas involucradas en el análisis de rendimiento de una red, aquí definiremos como primer elemento el “Índice de rendimiento” el cual si es representado por un valor pequeño tiene un rendimiento eficiente y si tiene un valor elevado tiene un rendimiento deficiente. El segundo elemento involucrado es el parámetro del espacio el cual ajusta los valores de pesos y bias de la red para mejorar el rendimiento.

En el capítulo investigaremos las características de las superficies de rendimiento y las condiciones para llegar a sus puntos mínimos y como lucen dichas superficies.

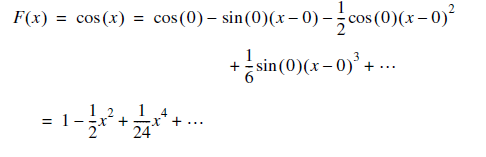
Debemos decir que el valor mínimo está representado por \bold{F(X)} donde x es un escalar de ajuste. Ahora asumamos que el índice de rendimiento es una función analítica, por lo tanto, su derivada existe. Por lo cual puede ser representada por la serie de Taylor expandida y relacionarla con un punto X\*



Usaremos la serie de Taylor para aproximarnos al valor optimo, limitando la expansión a un número finito de términos. Por ejemplo



La serie de Taylor F(x) es mínima cuando el punto X\*= 0 es



Graficando esto quedaría de la siguiente manera:

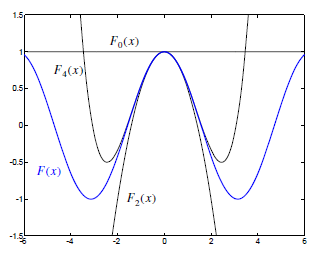


Figura 1: función coseno y las aproximaciones de la serie de Taylor

Podemos ver en la figura que las 3 aproximaciones son precisas si X\* = 0. Y si X se mueve más lejos de X\* solo la de mayor orden podrá aproximarse acertadamente. La de segundo orden es precisas con un rango más amplio que la de orden cero y la de cuarto orden es precisa solo con un valor más grande que con la de segundo orden. Una investigación de la ecuación de la figura explica este comportamiento. Cada término acertado en la serie involucra una mayor potencia de (x – x\*). Cuando x se acerca a x\*, estos términos se volverán geométricamente más pequeños.

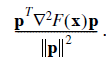
Usaremos las aproximaciones de la serie de Taylor para el índice de rendimiento e investigar su forma en el más acertado posible.

Derivadas direccionales:

Analíticamente podemos saber que sacando la derivada de cada elemento dentro de la ecuación de matrices de la serie de Taylor tendremos una función con direcciones arbitrarias. Por lo cual para solucionar esto tendremos que implementar algo llamador vector de dirección **p** el cual es un vector el cual tiene la dirección que queremos saber con la derivada. Esta derivada direccional puede ser computada de la siguiente manera:



Y la segunda derivada con recto a p puede ser computada por:



LA función tiene una inclinación cero en la dirección p desde el punto x\*, ¿Por qué sucedió? Que podemos decir acerca de esas direcciones que tienen una inclinación 0? Si consideramos la definición de la derivada direccional, podemos ver que el numerador es el producto punto entre la dirección del vector y el gradiente. Por lo cual cualquier dirección ortogonal con el gradiente será cero.

Qué dirección tiene la mejor dirección? La inclinación máxima ocurre cuando el producto punto de la dirección del vector y el gradiente es un máximo. Esto ocurre cuando la direccion del vector es la misma que el gradiente, este efecto es ilustrado en la figura 8.2 , que muestra una grafica de contornos y una 3d de F(x). En la gráfica de contornos vemos 5 vectores en nuestro punto nominal x\* apuntando a diferentes direcciones. La máxima derivada ocurre in la direccion del gradiente. La direccion cero esta en direccion ortogonal a el contorno (tangente a la linea del contorno)

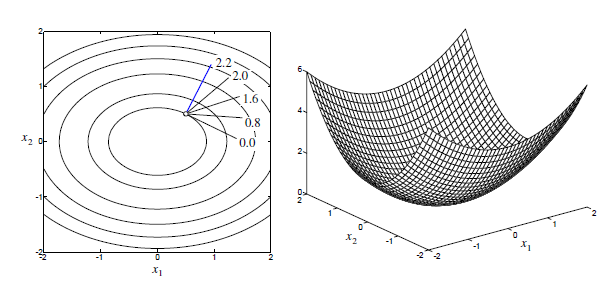
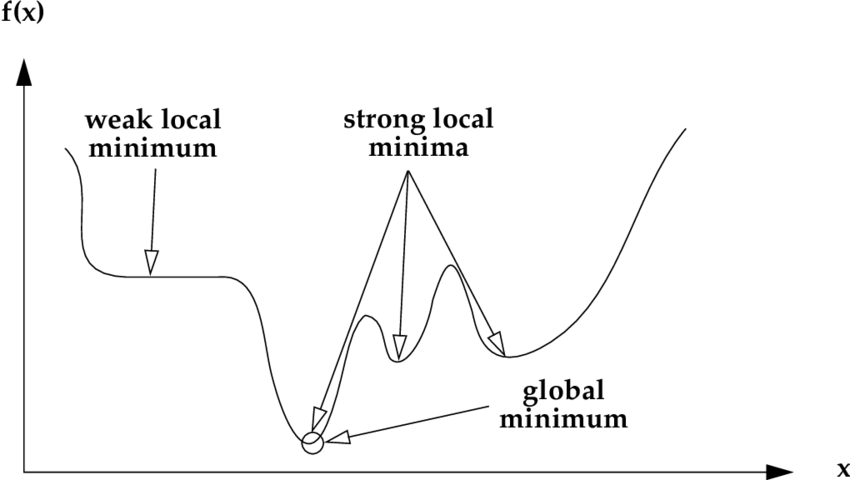


Figura 2: Función cuadrática y derivada direccionales

**mínimo:**

Ahora veamos las definiciones de los diferentes mínimos que puede haber en nuestra función en la figura 3:



**Figura 3: tipos de mínimo en un función F(x)**

Acorde a lo que nos muestra la Figura 3 podemos decir que:

Un fuerte mínimo local es aquel punto en la función en donde si existe un desplazamiento mayor a 0 habrá un incremento en cualquier dirección.

Un mínimo local flojo será aquel en donde si existe un desplazamiento mayor a 0 no habrá cambio en el valor de F(x).

Un mínimo global será aquel fuerte mínimo en donde no existe un punto en el plano que contenga un número mayor a F(x) en ese punto.

Definiendo estos conceptos podemos bajarlos a la representación en un plano y grafica de contornos como se muestran en la figura 4:

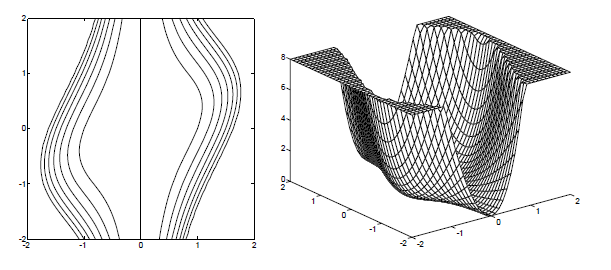


Figura 4: Grafico de contornos y plano representando un mínimo local flojo