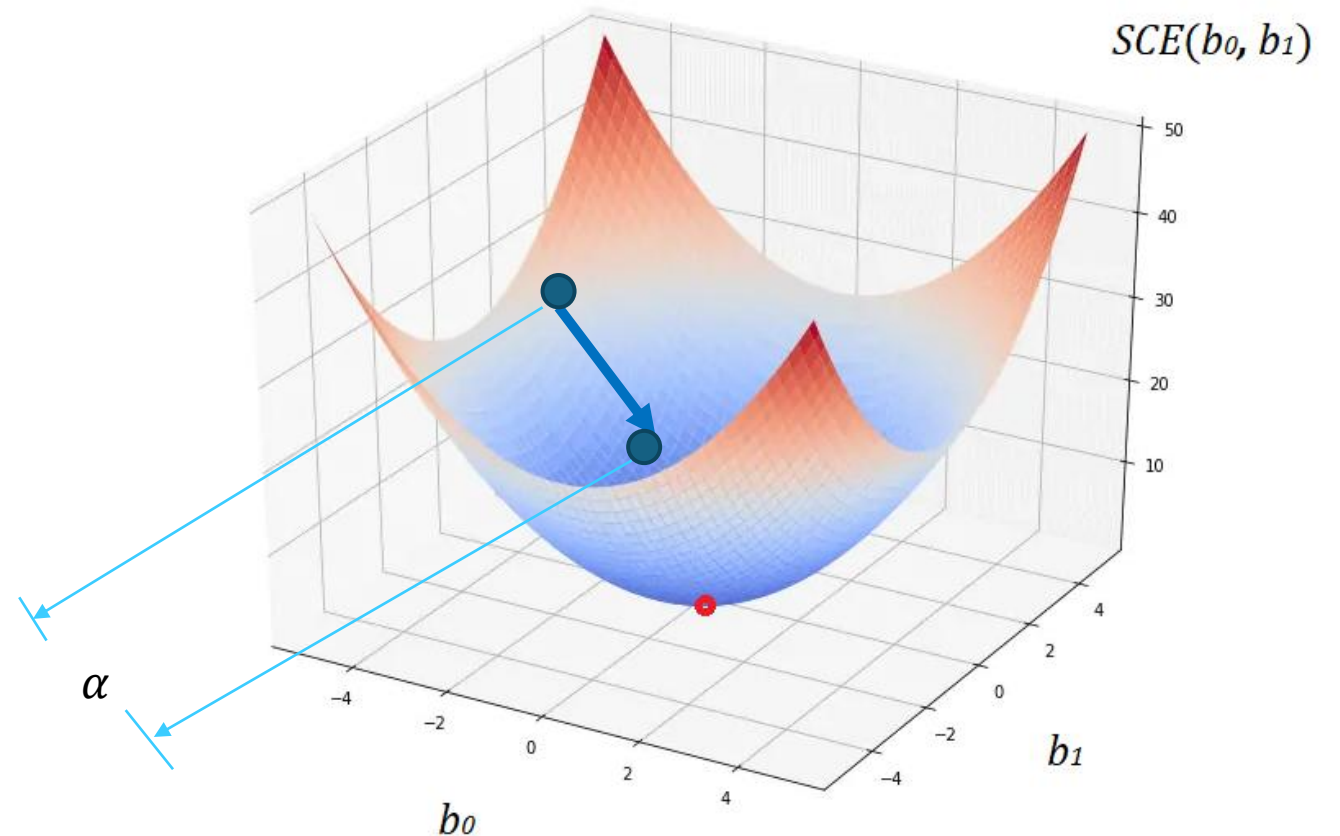


Regresión lineal con gradiente descendiente

Dra. Consuelo Varinia García Mendoza

Gradiente descendiente por lotes (BGD)

- Tamaño de paso = $\alpha \in (0,1)$
- Dirección \rightarrow derivada
- No. de pasos (iteraciones)



BGD

$$w_i = w_i - \alpha \frac{\partial f(w_i)}{\partial w_i}$$

$$w_i = w_i - 2\alpha \sum_{j=0}^{m-1} (w_i x_{j,i} - y_j) \cdot x_{j,i}$$

$$w_i = w_i - 2\alpha \cdot \text{sum} \left(\left(w_i \begin{bmatrix} x_{0,i} \\ x_{1,i} \\ \vdots \\ x_{m-1,i} \end{bmatrix} - \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{bmatrix} \right) \cdot \begin{bmatrix} x_{0,i} \\ x_{1,i} \\ \vdots \\ x_{m-1,i} \end{bmatrix} \right)$$

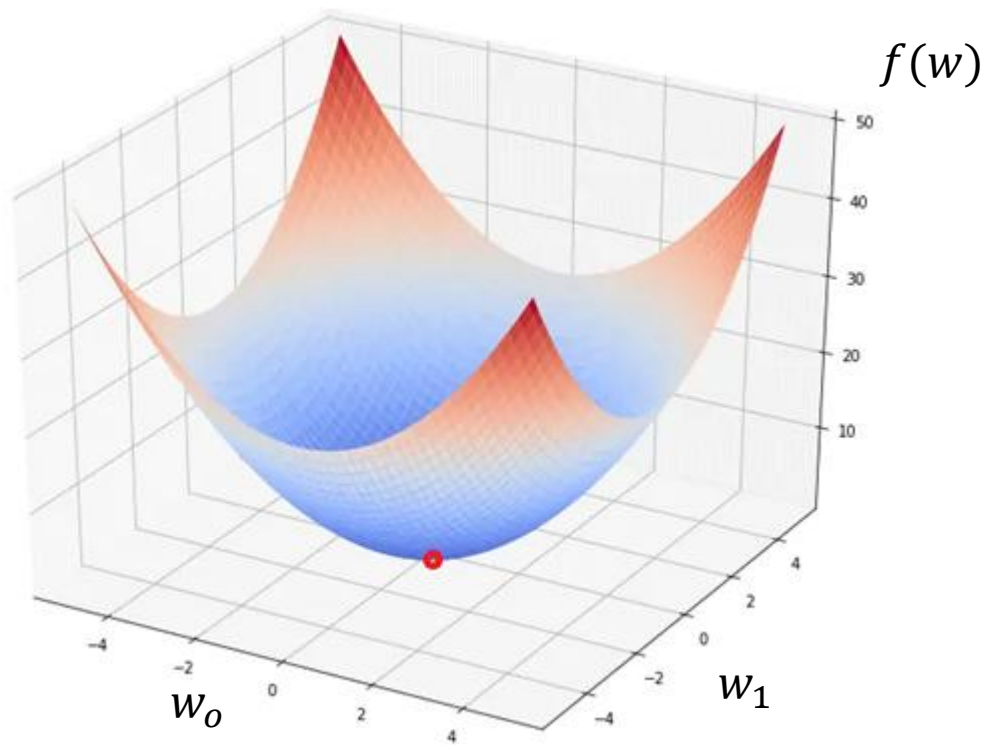
Limitaciones del algoritmo de BGD

- Permite ajustar los pesos de forma iterativa
- En cada iteración se calculan los pesos de cada característica y del error
- Considera todos los ejemplos en su cálculo
- Para conjuntos de datos grandes esto puede ser muy costoso y BGD se vuelve lento

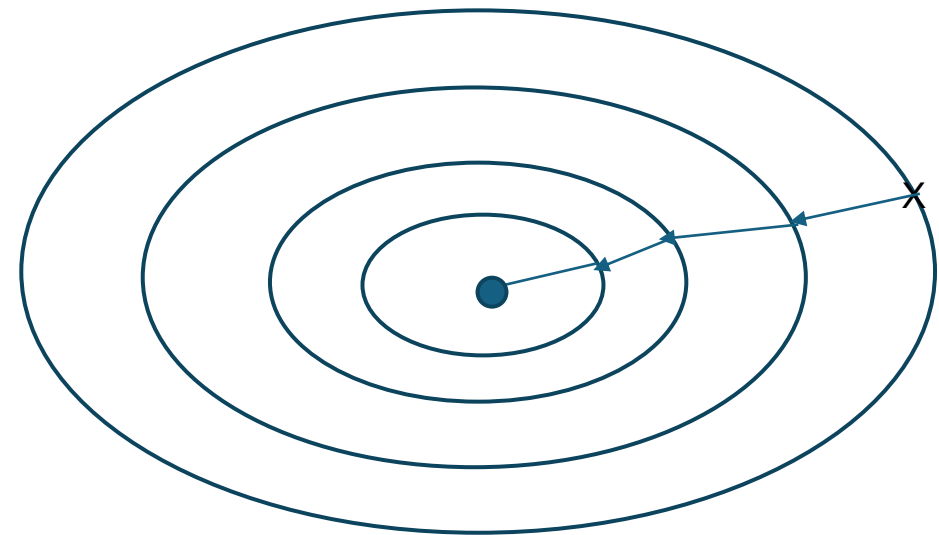
Gradiente descendente estocástico (SGD)

- Existe una variante del algoritmo BGD que considera sólo un ejemplo para el ajuste de pesos en cada iteración. Eligiendo de manera aleatoria este ejemplo es decir se establece r de manera aleatoria en cada iteración
- Con este único ejemplo se ajustan los pesos y se intenta una aproximación al óptimo
- Este algoritmo reduce el número de operaciones que se realizan en cada iteración por lo que puede manejar grandes cantidades de datos
- Sin embargo, se debe considerar que no siempre se encontrarán los pesos óptimos, pero si una buena aproximación

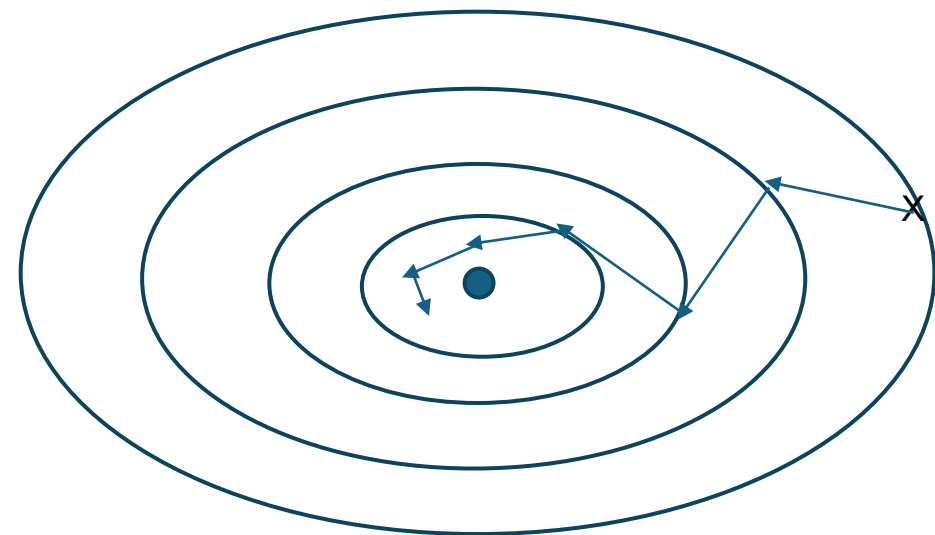
$$w_k = w_k - 2\alpha(w_k x_{r,k} - y_r) \cdot x_{r,k}$$



BGD



SGD



Implementación de BGD o SGD

- Funciones de perdida
 - Suma de errores al cuadrado

$$f(w_k) = \sum_{r=0}^{e-1} (w_k x_{r,k} - y_r)^2$$

- Error cuadrático medio

$$f(w_k) = \frac{1}{e} \sum_{r=0}^{e-1} (w_k x_{r,k} - y_r)^2$$

Regresión lineal

- Mínimos Cuadrados Ordinarios (OLS)

- Método analítico

$$\frac{\partial \text{SCE}(\mathbf{W})}{\partial w_k} = 0$$

- Gradiente descendiente por lotes (BGD)

- $w_k = w_k - 2\alpha \sum_{r=0}^{e-1} (w_k x_{r,k} - y_r) \cdot x_{r,k}$

- Gradiente descendiente estocástico (SGD)

- $w_k = w_k - 2\alpha (w_k x_{r,k} - y_r) \cdot x_{r,k}$

Aprendizaje supervisado