Regresión Lineal + Gradiente Descendente

Regresión lineal (mínimos cuadrados)

El modelo de regresión lineal que se puede usar es el siguiente:

$$\mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

asumiendo que el bias está incluido en $\mathbf{w}$, y la matriz de diseño tiene un vector adicional. Es posible resolver este modelo usando *ecuaciones normales*, i.e., solución analítica.

Regresión lineal (mínimos cuadrados)
Fuerza bruta

Una forma de ajustar un modelo lineal (y cualquier red neuronal) es inicializando los parámetros en 0's o en valores aleatorios muy pequeños. De esta forma, en k iteraciones:

- Elegir otro conjunto aleatorio de pesos.
- Si el desempeño del modelo es mejor, se conservan los nuevos pesos.
- Si el desempeño del modelo es peor, se descartan.

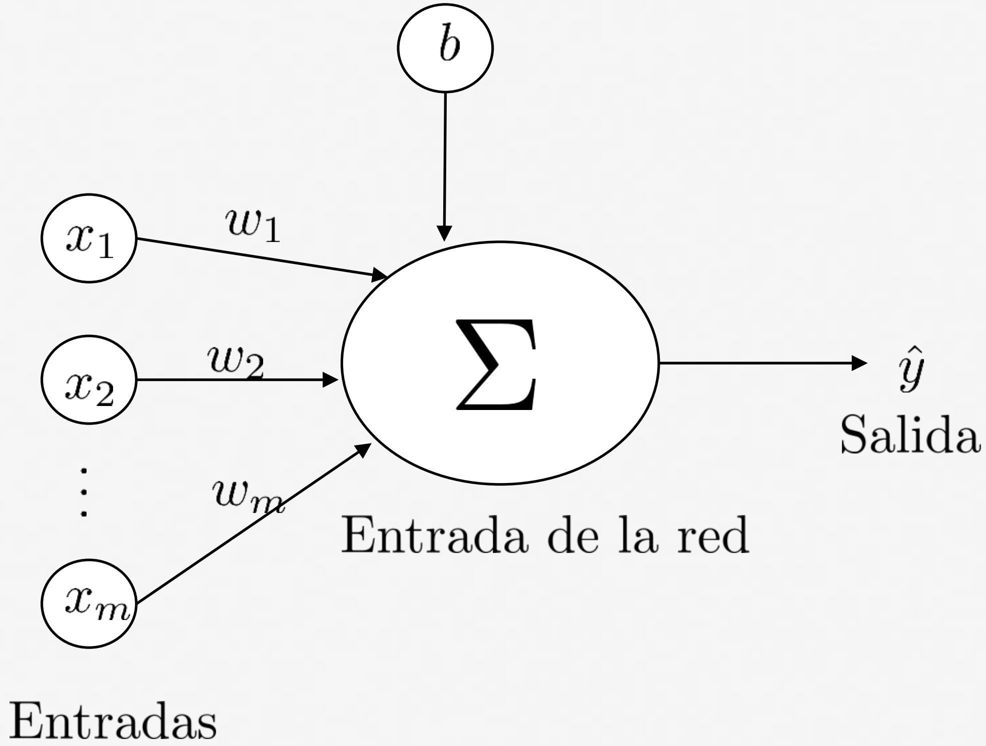
Este método asegura encontrar la solución óptima para un k bastante grande, pero sería demasiado lento.

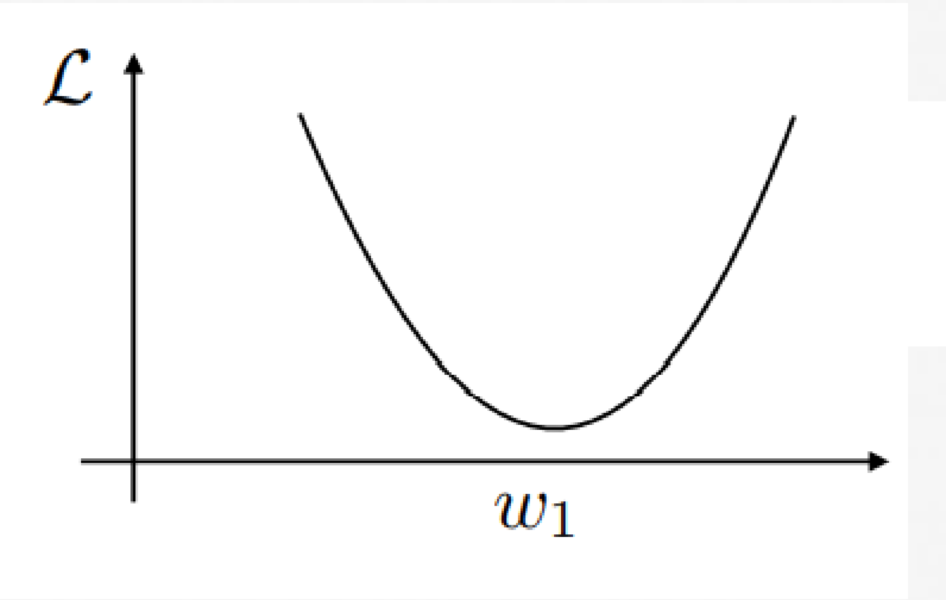
Regresión lineal (mínimos cuadrados)
Una mejor manera!

Una mejor forma de hacerlo, es analizando qué efecto tiene un cambio en un parámetro en el desempeño preditivo (pérdida) del modelo; posteriormente, se cambiará solo un poco el peso en la dirección que mejora el desempeño (minimiza la pérdida). Esto se hace con bastantes pasos pequeños, hasta que la pérdida no se decremente más.

Institución Universitaria
Reacreditada en Alta Calidad

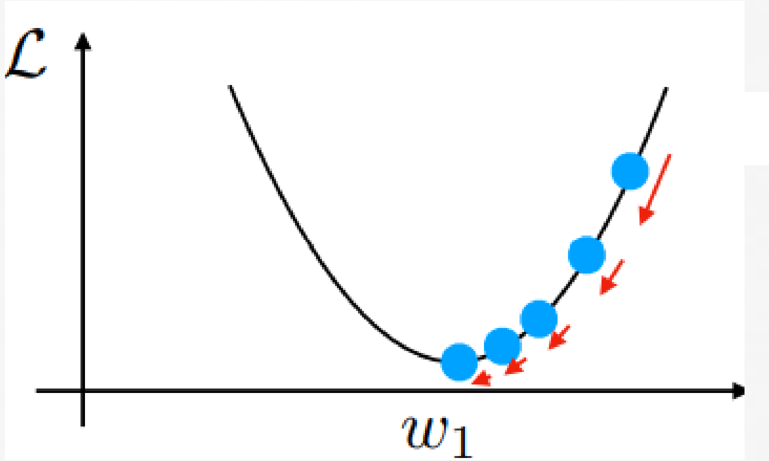
Entrenamiento del regresor lineal con gradiente descendente

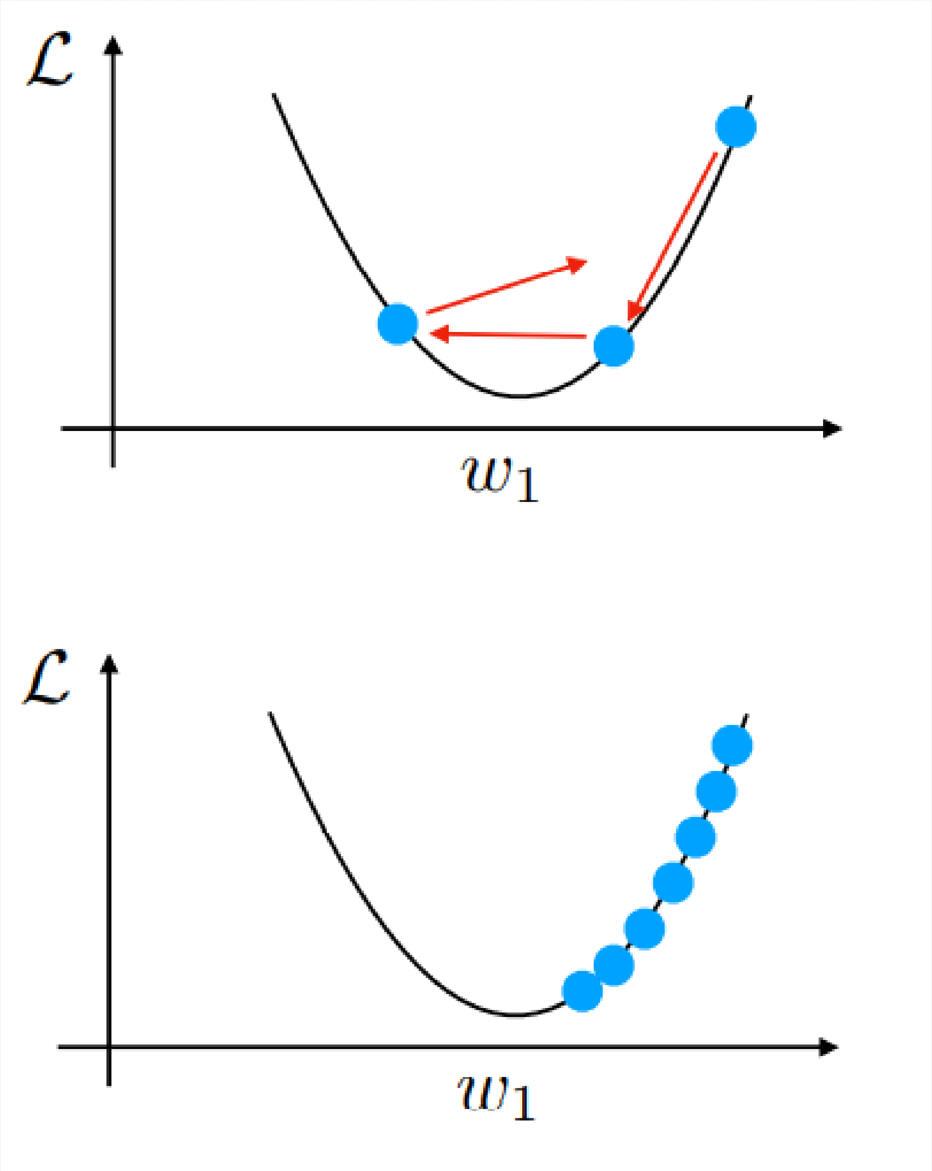




Función de pérdida convexa

$$\mathcal{L}(\mathbf{w}, b) = \sum_i (y^{[i]} - \hat{y}^{[i]})^2$$


$\mathcal{L}$
w_1


$\mathcal{L}$
w_1
$\mathcal{L}$
w_1

Institución Universitaria
Reacreditada en Alta Calidad

Derivada de la función de pérdida para regresión lineal

Dada la función de pérdida

$$\mathcal{L}(\mathbf{w}, b) = \sum_i (y^{[i]} - \hat{y}^{[i]})^2,$$

la derivada estaría dada por

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_j} &= \frac{\partial}{\partial w_j} \sum_i (y^{[i]} - \hat{y}^{[i]})^2 \\
&= \frac{\partial}{\partial w_j} \sum_i (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]})^2 \\
&= \sum_i 2(y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \frac{\partial}{\partial w_j} (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \\
&= \sum_i 2(y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \frac{\partial}{\partial w_j} (-\mathbf{w}^\top \mathbf{x}^{[i]}) \\
&= \sum_i 2(\mathbf{w}^\top \mathbf{x}^{[i]} - y^{[i]}) x_j^{[i]}
\end{aligned}$$

Derivada de la función de pérdida para regresión lineal: caso alternativo

A menudo, la función de pérdida se escala por un factor de $\dfrac{1}{2n}$ por conveniencia:

$$\mathcal{L}(\mathbf{w}, b) = \tfrac{1}{2n} \textstyle\sum_i (y^{[i]} - \hat{y}^{[i]})^2.$$

La derivada estaría dada por

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_j} &= \frac{\partial}{\partial w_j} \frac{1}{2n} \textstyle\sum_i (y^{[i]} - \hat{y}^{[i]})^2 \\
&= \frac{\partial}{\partial w_j} \textstyle\sum_i \frac{1}{2n} (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]})^2 \\
&= \textstyle\sum_i \frac{1}{n} (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \frac{\partial}{\partial w_j} (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \\
&= \textstyle\sum_i \frac{1}{n} (y^{[i]} - \mathbf{w}^\top \mathbf{x}^{[i]}) \frac{\partial}{\partial w_j} (-\mathbf{w}^\top \mathbf{x}^{[i]}) \\
&= \textstyle\sum_i \frac{1}{n} (\mathbf{w}^\top \mathbf{x}^{[i]} - y^{[i]}) x_j^{[i]}
\end{aligned}$$

Institución Universitaria
Reacreditada en Alta Calidad
80 Años

¡MUCHAS GRACIAS!

VIGILADA MINEDUCACIÓN

www.itm.edu.co

Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación