Regresión Lineal: Derivación Analítica y Gradiente Descendente
(Hojas 04–07)

1. Función de costo en forma vectorial

La función de costo para regresión lineal puede escribirse como:

$$J(\boldsymbol{\theta}) = \frac{1}{2N} \left\| \boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y} \right\|^2$$

donde $\boldsymbol{X} \in \mathbb{R}^{N \times d}$ es la matriz de diseño (con columna de unos si hay intercepto) y $\boldsymbol{y}$ es el vector de respuestas.

2. Derivada del costo

Usando derivadas matriciales:

$$J(\boldsymbol{\theta}) = \frac{1}{2N}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^\top(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$$

El gradiente es:

$$\nabla_{\boldsymbol{\theta}} J = \frac{1}{N}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$$

3. Solución Analítica (Mínimos Cuadrados)

Buscando el mínimo de J, se iguala el gradiente a cero:

$$\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}) = 0$$

$$\Rightarrow \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{X}^\top\boldsymbol{y} \quad \Rightarrow \quad \hat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

Esta es la solución de mínimos cuadrados ordinarios (OLS), válida cuando $\boldsymbol{X}^\top\boldsymbol{X}$ es invertible.

4. Descenso por Gradiente

En lugar de resolver analíticamente, podemos aplicar un algoritmo iterativo:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \alpha \cdot \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}^{(t)})$$

Sustituyendo el gradiente:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \alpha \cdot \frac{1}{N} \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\theta}^{(t)} - \boldsymbol{y})$$

donde α es la tasa de aprendizaje.

5. Interpretación geométrica

La función de costo $J(\boldsymbol{\theta})$ es convexa, con forma de parábola en el caso univariado y de hiperparaboloide en el caso multivariado. Por tanto, cualquier mínimo local es también global.

6. Derivación de la actualización paso a paso

Para cada paso del gradiente descendente:

$$\begin{aligned}
\text{Error:} \quad & \boldsymbol{e}^{(t)} = \boldsymbol{X}\boldsymbol{\theta}^{(t)} - \boldsymbol{y} \\
\text{Gradiente:} \quad & \boldsymbol{g}^{(t)} = \frac{1}{N} \boldsymbol{X}^\top \boldsymbol{e}^{(t)} \\
\text{Actualización:} \quad & \boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \alpha \boldsymbol{g}^{(t)}
\end{aligned}$$

7. Elección de la tasa de aprendizaje

Si α es muy grande, puede que no converja. Si es muy pequeño, puede tardar demasiado en converger. Se debe usar una validación cruzada o ensayo-error para ajustarlo.