5.2 Regola del gradiente per la regressione logistica

La regressione logistica può essere ottenuta considerando come funzione di costo la *cross-entropia*

$$\ell(\hat{y}, y) = -y \log \hat{y} - (1-y)\log(1-\hat{y}),$$

dove $\hat{y} = h(x) = \sigma(w^\top x)$ (si ricordi che $w \in \mathbb{R}^{d+1}$ è un vettore di parametri e σ denota la funzione sigmoide logistica).

Teorema 5.2.1. *La regola di aggiornamento del gradiente per la regressione logistica è*

$$w \leftarrow w - \eta(h(x) - y)x.$$

Dimostrazione. Per concretizzare la regola di aggiornamento del gradiente nell'algoritmo della discesa stocastica del gradiente

$$w \leftarrow w - \eta \nabla \ell,$$

calcoliamo le derivate parziali $\partial \ell / \partial w_j$:

$$\begin{aligned}
\frac{\partial \ell}{\partial w_j} &= -y \frac{\partial}{\partial w_j} \log \hat{y} - (1-y) \frac{\partial}{\partial w_j} \log(1-\hat{y}) \\
&= -y \frac{1}{\hat{y}} \sigma(z)(1-\sigma(z)) \frac{\partial}{\partial w_j} z + (1-y) \frac{1}{1-\hat{y}} \sigma(z)(1-\sigma(z)) \frac{\partial}{\partial w_j} z \\
&= -y \frac{1}{\hat{y}} \hat{y}(1-\hat{y}) x_j + (1-y) \frac{1}{1-\hat{y}} (1-\hat{y}) \hat{y} x_j \\
&= (-y(1-\hat{y}) + (1-y)\hat{y})\, x_j \\
&= (\hat{y} - y)\, x_j \\
&= (h(x) - y)\, x_j. \qquad \square
\end{aligned}$$