5.1 Livello di incertezza delle predizioni

In generale, il livello di incertezza associabile alla predizione su un punto di input $x \in \mathcal{X}$ è quantificabile con la formula $\mathbb{E}_{y|x}\, \ell(h(x), y)$.

5.1.1 Regressione

Nella regressione con costo quadratico, si può dimostrare che il livello di incertezza su x è dato semplicemente dalla varianza condizionata $\mathrm{Var}(y|x)$. Se la regressione è lineare, tale varianza condizionata coincide con la varianza degli scarti $\epsilon = y - h(x)$ (che per ipotesi esplicita od implicita hanno distribuzione normale).

5.1.2 Classificazione

Nella classificazione basata su stime di probabilità, l'incertezza sulla predizione per x è

$$\mathbb{E}_{y|x}\, \ell(h(x), y) = \Pr_y(h(x) \neq y|x).$$

In altre parole, il livello di *certezza* di una predizione $h(x) = j$ è quantificabile grazie alla probabilità condizionata

$$\Pr(y = j|x) = \frac{\Pr(x, y = j)}{\Pr(x)} = \frac{\Pr(x|y = j)\Pr(y = j)}{\sum_{i=1}^{K} \Pr(x|y = i)\Pr(y = i)}.$$

Utilizzando la notazione $P_i(x) \stackrel{\text{def}}{=} \Pr(x|y = i)$, $\pi_i \stackrel{\text{def}}{=} \Pr(y = i)$,

$$\Pr(y = j|x) = \frac{\pi_j P_j(x)}{\sum_{i=1}^{K} \pi_i P_i(x)}.$$

Consideriamo in particolare problemi di classificazione binaria basata su probabilità: $\mathcal{Y}_0 = \{0, 1\}$, $\mathcal{Y} = [0, 1]$. Possiamo scrivere

$$\begin{aligned}\Pr(y = 1|x) = \frac{\pi_1 P_1(x)}{\pi_0 P_0(x) + \pi_1 P_1(x)} &= \frac{1}{1 + \frac{\pi_0 P_0(x)}{\pi_1 P_1(x)}} \\ &= \frac{1}{1 + \exp(-z)}\end{aligned}$$

dove abbiamo introdotto una nuova funzione $z(x) \stackrel{\text{def}}{=} \log(\pi_1 P_1(x)/\pi_0 P_0(x))$. Ricordando che il discriminante della classe j è $\delta_j(x) = \log(\pi_j P_j(x))$, si noti che $z(x) = \delta_1(x) - \delta_0(x)$ è positiva quando la predizione è 1, e negativa quando la predizione è 0.

La funzione $z \mapsto \sigma(z) = 1/(1 + \exp(-z))$ è chiamata *sigmoide logistica*, o semplicemente *sigmoide*.