

Indice

1	Metodi lineari di classificazione	2
1.1	Introduzione	2

Capitolo 1

Metodi lineari di classificazione

1.1 Introduzione

In questo capitolo rivisitiamo il problema di classificazione e ci concentriamo in metodi lineari di classificazione. Dato che il nostro predittore $G(x)$ prende valori di un insieme discreto \mathcal{G} , possiamo sempre dividere lo spazio di input in una collezione di regioni secondo la classificazione. Abbiamo visto nel Capitolo 2 che le frontiere di questi regioni possono essere ruvide o lisce, dipendendo della funzione di predizione. Per una classe importante di procedure, queste *frontiere di decisione* sono lineari; questo è ciò che intendiamo con metodi lineari di classificazione.

Ci sono diversi modi in cui le frontiere di decisione lineari possono essere trovate. Nel Capitolo 2 abbiamo adattato modelli di regressione lineare alle variabili indicatori di classi, e dopo classifichiamo secondo il **largest fit**. Supponiamo che ci sono K classi, etichettate come $1, 2, \dots, K$ per comodità, ed il modello lineare adattato per la k -esima variabile indicatore di risposta sia $\hat{f}_k(x) = \hat{\beta}_{k0} + \hat{\beta}_k^T x$. La frontiera di decisione tra la classe k e l è il insieme di punti per cui $\hat{f}_k(x) = \hat{f}_l(x)$, cioè, il insieme $\{x : (\hat{\beta}_{k0} - \hat{\beta}_{l0}) + (\hat{\beta}_k - \hat{\beta}_l)^T x = 0\}$, che è un insieme affine o un iperpiano¹. Siccome lo stesso è vero per ogni paio di classi, lo spazio di input è diviso in regioni di classificazione costante, con frontiere di decisione che sono iperplanare a tratti. Questo approccio con regressione fa parte di una classe di metodi che modellano *funzioni discriminanti* $\delta_k(x)$ per ogni classe, e dopo classificano x alla classe con il valore più grande nella sua funzione discriminante. Metodi che modellano le probabilità posteriori $\Pr(G = k|X = x)$ appartengono anche a questa classe. Chiaramente, se $\delta_k(x)$ o $\Pr(G = k|X = x)$ sono lineare in x , allora le frontiere di decisione saranno anche lineare.

In realtà, tutto ciò di cui abbiamo bisogno è che alcuna trasformazione monotona di δ_k o $\Pr(G = k|X = x)$ sia lineare affinché le frontiere di decisione siano lineari. Per esempio, se ci

¹In senso stretto, un iperpiano attraversa l'origine, mentre un insieme affine non necessariamente. A volte ignoriamo la distinzione e intendiamo iperpiani in generale

sono due classi, un modello popolare per le probabilità posteriori è

$$\begin{aligned}\Pr(G = 1|X = 1) &= \frac{e^{\beta_0 + \beta^T x}}{1 + e^{\beta_0 + \beta^T x}}, \\ \Pr(G = 2|X = 1) &= \frac{1}{1 + e^{\beta_0 + \beta^T x}}.\end{aligned}\tag{1.1}$$

La trasformazione monotona qui è la trasformazione *logit*: $\ln[p/(1-p)]^2$, e infatti vediamo che

$$\ln \frac{\Pr(G = 1|X = 1)}{\Pr(G = 2|X = 1)} = \beta_0 + \beta^T x \tag{1.2}$$

²Se $p_1 = \Pr(G = 1|X = 1)$ e $p_2 = \Pr(G = 2|X = 1)$, poi abbiamo che $\ln[p_1/(1-p_1)] = \beta_0 + \beta^T x$ e $\ln[p_2/(1-p_2)] = -(\beta_0 + \beta^T x)$.