

# Model probabilístic generatiu per variables binàries

Albert Ribes

10 de diciembre de 2017

Considerem el cas de tenir  $d$  variables binàries  $x_i \in \{0, 1\}$  en un problema de classificació en  $K$  classes,  $C_1, \dots, C_K$ . La distribució conjunta  $P(x) = P(x_1, \dots, x_d)$  requereix en principi el coneixement de  $2^d - 1$  números (les respectives probabilitats de cada combinació) per cada classe, la qual cosa no és factible. Decidim doncs treballar amb distribucions condicionals (per cada classe  $k$ ) de la forma:

$$P(x|C_k) = \prod_{i=1}^d p_{k_i}^{x_i} (1 - p_{k_i})^{1-x_i}$$

on  $p_{k_i}$  és la probabilitat de tenir un 1 a la variable binària  $i$  per la classe  $k$ , que es pot estimar de les dades. Es demana:

1. Argumenteu per què aquesta decissió correspon a assumir que les  $d$  variables binàries són estadísticament independents donada la classe.
2. Doneu l'expressió per les funcions discriminants  $a_k(x)$  que en resulten. Són discriminants lineals?
3. Doneu l'expressió per la probabilitat a posteriori  $P(C_k|x)$ .