



Machine Learning 1


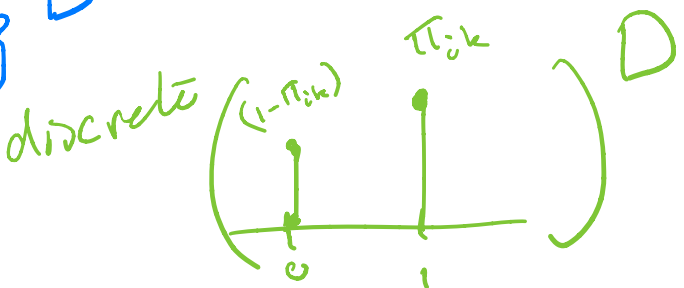
Lecture 6.2 - Supervised Learning
Classification - Probabilistic Generative
Models - For Discrete Variables

Erik Bekkers

(Bishop 4.2.3)



Probabilistic Generative Models: Discrete

- Input: discrete feature vectors $\mathbf{x}_n = (x_1, \dots, x_D)^T$
 $x_i \in \{0, 1\}$
 $x_n \in \{0, 1\}^D$
 $x_n = (0, 1, 1, 0, \dots, 1)^T$
- For D-dimensional input:


of parameters for $p(\mathbf{x}|C_k)$ per class: $2^D - 1$

- Naive Bayes assumption:** feature values are treated as independent when conditioned on class C_k !
 $\pi_{ki} = p(x_i = 1 | C_k)$

- $$p(\mathbf{x}|C_k) = \prod_{i=1}^D p(x_i | C_k) = \prod_{i=1}^D \pi_{ki}^{x_i} (1 - \pi_{ki})^{1-x_i}$$
- $$p(C_k | \mathbf{x}) = \frac{\exp(a_k(\mathbf{x}))}{\sum_{j=1}^K \exp(a_j(\mathbf{x}))}$$

params per class = D

$$a_k = \ln p(\mathbf{x} | C_k) p(C_k) = \ln p(\mathbf{x} | C_k) + \ln p(C_k)$$

$$= \sum_{i=1}^D \{x_i \ln \pi_{ki} + (1-x_i) \ln (1 - \pi_{ki})\} + \ln p(C_k)$$

↑ via ML