

Exercise: estimate prior probabilities from training data

Example	Positive mammogram	Family history	alcohol	Cancer
1	Yes	Yes	Yes	yes
2	Yes	Yes	No	Yes
3	No	Yes	yes	Yes
4	Yes	No	No	No
5	Yes	No	Yes	No
6	No	No	Yes	No
7	No	No	No	No

Given the above training data set:

Calculate prior probability for each class: $P(C) = N_c/N$

$P(\text{cancer}=\text{yes})=?$

$P(\text{cancer}=\text{no})=?$

Exercise: estimate conditional probabilities from training data

Then calculate the conditional probabilities for each attribute

$$P(A_i | C_k) = |A_{ik}| / N_c \text{ E.g.}$$

$$P(\text{pos_mammo}=\text{yes} | \text{cancer}=\text{yes}) = ?$$

$$P(\text{pos_mammo}=\text{no} | \text{cancer}=\text{yes}) = ?$$

$$P(\text{pos_mammo}=\text{yes} | \text{cancer}=\text{no}) = ?$$

$$P(\text{pos_mammo}=\text{no} | \text{cancer}=\text{no}) = ?$$

Exercise: estimate conditional probabilities from training data

Repeat for the “family_history” attribute:

$$P(\text{family_history}=\text{yes} \mid \text{cancer}=\text{yes}) = ?$$

$$P(\text{family_history}=\text{no} \mid \text{cancer}=\text{yes}) = ?$$

$$P(\text{family_history}=\text{yes} \mid \text{cancer}=\text{no}) = ?$$

$$P(\text{family_history}=\text{no} \mid \text{cancer}=\text{no}) = ?$$

Exercise: estimate conditional probabilities from training data

Repeat for the “Alcohol” attribute:

$$P(\text{Alcohol} = \text{yes} \mid \text{cancer} = \text{yes}) = ?$$

$$P(\text{Alcohol} = \text{no} \mid \text{cancer} = \text{yes}) = ?$$

$$P(\text{Alcohol} = \text{yes} \mid \text{cancer} = \text{no}) = ?$$

$$P(\text{Alcohol} = \text{no} \mid \text{cancer} = \text{no}) = ?$$

Exercise: calculate posterior probabilities

Given a test case (pos_mammo=yes, fam_hist=yes, alcohol=yes), what is the prediction, cancer or no cancer?

$P(\text{cancer} \mid \text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$

vs.

$P(\text{no cancer} \mid \text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$

Exercise: calculate posterior probabilities

- $P(\text{cancer} \mid \text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$
= $P(\text{pos_mammo}=\text{yes} \mid \text{cancer}) * P(\text{fam_hist}=\text{yes} \mid \text{cancer}) * P(\text{alcohol}=\text{yes} \mid \text{cancer}) * P(\text{cancer}) / P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$
= $(2/3 * 3/3 * 2/3 * 3/7) / P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$
= $(12/63) / P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes})$

Exercise: calculate posterior probabilities

- Similarly,
$$\begin{aligned} & P(\text{no cancer} \mid \text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes}) \\ &= P(\text{pos_mammo}=\text{yes} \mid \text{no cancer}) * \\ &\quad P(\text{fam_hist}=\text{yes} \mid \text{no cancer}) * \\ &\quad P(\text{alcohol}=\text{yes} \mid \text{no cancer}) * \\ &\quad P(\text{no cancer}) / \\ &\quad P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes}) \\ &= (2/4 * 0/4 * 2/4 * 4/7) / \\ &\quad P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes}) \\ &= (0) / \\ &\quad P(\text{pos_mammo}=\text{yes}, \text{fam_hist}=\text{yes}, \text{alcohol}=\text{yes}) \end{aligned}$$

Exercise: calculate posterior probabilities

- Prediction: **cancer!**
 - 12/63 : 0
 - Choose the decision with max posterior probability