Data Science 8

TILBURG UNIVERSITY

Understanding Society

Overview

- Bayesian learning
- Bayes' Rule
- Naive Bayes classifier

Bayesian Learning: Key Concepts

- Data (d)
- Hypotheses (h_i)
- Evidence

- Bayes' rule

$$P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)$$

Bayes in Action

$$P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)$$

posterior probability

likelihood

hypothesis prior

New observations update your beliefs

$$P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)$$

posterior probability

likelihood

hypothesis prior

- You encounter a person with long hair and try to determine the probability of the person being female

- prior = 0.5 (both genders are equally likely)
- likelihood = P(long hair | female)

- Posterior probability updates the future prior (e.g., flower power)

Example (adapted from Russell & Norvig)

Five types of candy bags

h_1: 100% cherry
h_2: 75% cherry + 25% lime
h_3: 50% cherry + 50% lime
h_4: 25% cherry + 75% lime
h_5: 100% lime

Which one is it?

?

Candy in two flavors

Bayesian prediction

- Given a new opaque bag of candy
- H denotes the type of bag (h1,h2,h3,h4,h5)
- **d** denotes all observations of cherry and lime
- TASK: predict the flavour of the next piece of candy

- In Bayesian learning: Calculate the probability of each hypothesis, given the observations (data)

$$\boxed{P(h_i|\mathbf{d}) = \alpha P(\mathbf{d}|h_i)P(h_i)}$$

Calculating Likelihoods

h_1: 100% cherry
h_2: 75% cherry + 25% lime
h_3: 50% cherry + 50% lime
h_4: 25% cherry + 75% lime
h_5: 100% lime

- With each observation, the likelihood is computed according to

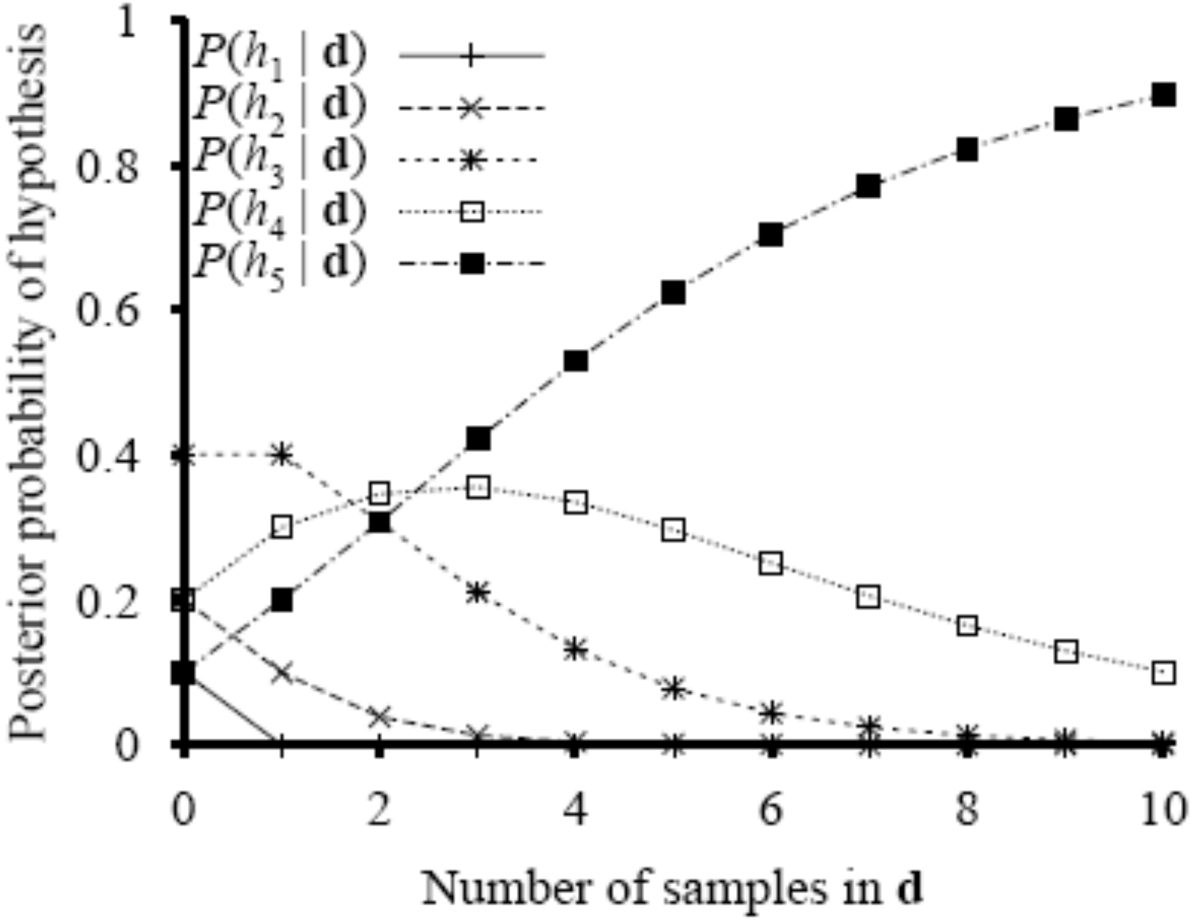
  P(d|h) = P(cherry|h) x P(lime|h) <— assuming independence

- In case the bag is all lime (h5) and the first 10 observations are lime, than

  P(d|h3) = 0.5^{10} and P(d|h5) = 1^{10}

TILBURG UNIVERSITY

Development of the posterior probabilities (h5 is the true hypothesis, all observations are 'lime')



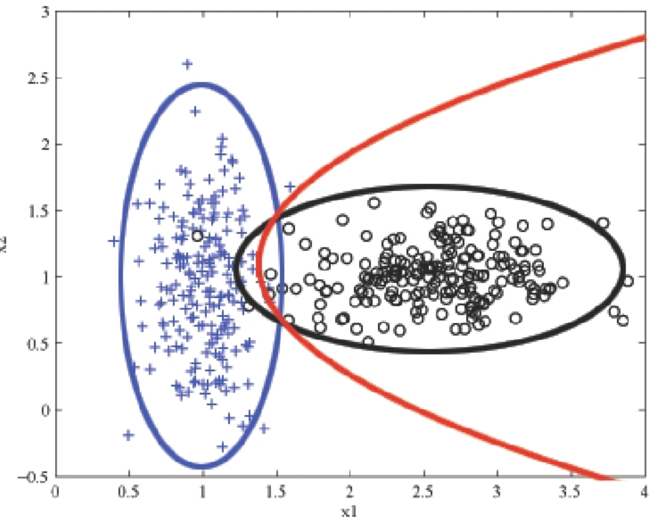
Naive Bayes Classifier

- Given a dataset with two classes (Dem,Rep) and three features (A,B,C), compute

  P(Dem|data) = P(Dem) P(A|Dem) x P(B|Dem) x P(C|Dem)
  P(Rep|data) = P(Rep) P(A|Rep) x P(B|Rep) x P(C|Rep)

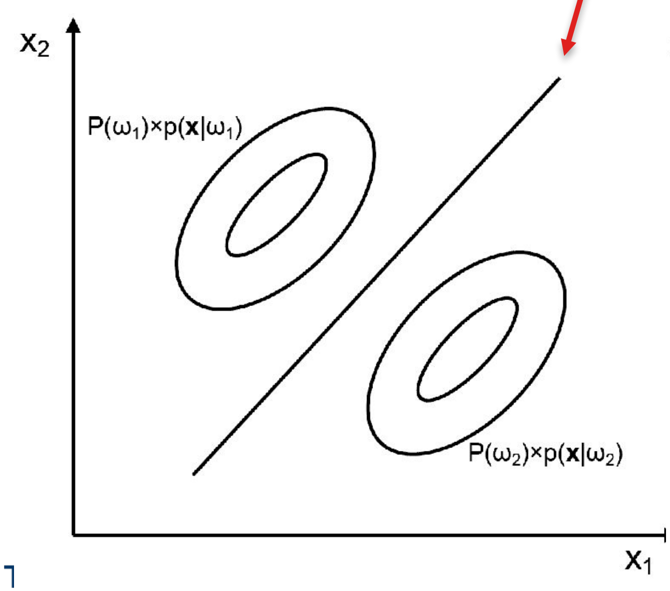
- IF P(Dem|data) > P(Rep|data)
  THEN Classification is Dem
  ELSE Classification is Rep

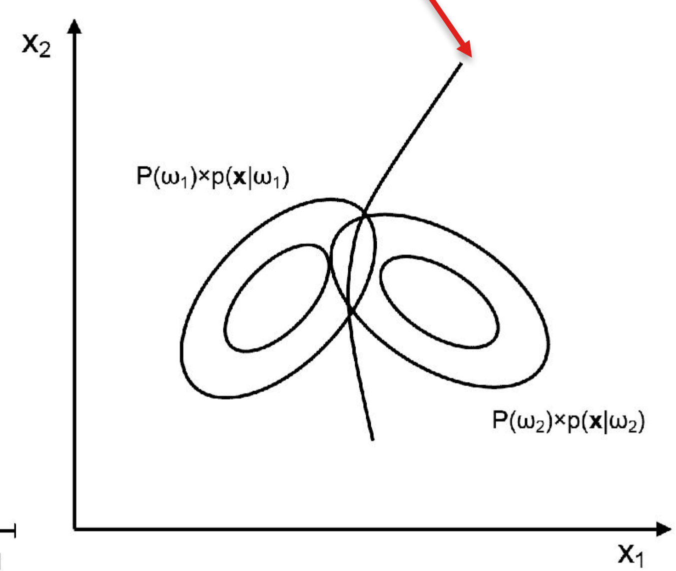
Naive Bayes decision boundaries



probabilities for black and blue classes are equal

can be straight or curved





TILBURG UNIVERSITY

ORANGE

- Use VOTING.TAB