Bayesian Statistics Project

Oscar Arandes Tejerina
Department of Physics, Stockholm University, AlbaNova University Center, 10691 Stockholm, Sweden

(Dated: January 16, 2026)

This is for the home project in the course *Bayesian Statistics for Machine Learning 3.0 DT8057* given by Halmstad University.

TABLE I. Description of Heart Disease Dataset Columns

Feature	**Description**
age	Age of the patient (in years).
sex	Gender of the patient (‘`male` → 1’, ‘`female`→ 0’).
chest-pain	Type of chest pain experienced: ‘`abnang` → 0’ (atypical angina), ‘`angina` → 1’ (typical angina), ‘`asympt` → 2’ (asymptomatic), ‘`notang` → 3’ (non-anginal pain).
resting-blood-pressure	Resting blood pressure in mm Hg.
serum-cholestoral	Serum cholesterol in mg/dl.
fasting-blood-sugar	Fasting blood sugar > 120 mg/dl: ‘`1` = true’, ‘`0` = false’.
rest-ecg	Resting electrocardiographic results: ‘`abn` → 0’ (ST-T wave abnormality), ‘`hyp` → 1’ (probable or definite left ventricular hypertrophy). ‘`norm` → 2’ (normal).
max-heart-rate	Maximum heart rate achieved during exercise.
exercise-angina	Exercise-induced angina: ‘`1` = yes’, ‘`0` = no’.
oldpeak-ecg	ST depression induced by exercise relative to rest.
slope-ecg	Slope of the peak exercise ST segment: ’`down` → 0’ (downsloping), ‘`flat` → 1’ (flat), ‘`up` → 2’ (upsloping).
major-vessels	Number of major vessels (0–3) colored by fluoroscopy.
thal	Thalassemia: ‘`fix` → 0’ (fixed defect), ‘`norm` → 1’ (normal), ‘`rev` → 2’ (reversible defect).
H, S1, S2, S3, S	One-hot encoding of the output class (disease category). Example: ‘H=0, S1=0, S2=1, S3=0, S=0‘ indicates class ‘S2‘.

NAIVE BAYES CLASSIFIER FOR MIXED DATA

Given a dataset with class variable $Y \in \{1, \ldots, k\}$ and feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, the posterior probability is

$$P(Y = y \mid \mathbf{x}) \propto P(\mathbf{x} \mid Y = y)P(Y = y) = \prod_{i=1}^{d} P(x_i \mid Y = y)P(Y = y)$$

The Naive Bayes classifier assigns a new instance to the class

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} \left[P(Y = y) \prod_{i=1}^{d} P(x_i \mid Y = y) \right].$$

$$\hat{y} = \arg\max_{y \in \mathcal{Y}} \left[\log P(Y = y) + \sum_{i=1}^{d} \log P(x_i \mid Y = y) \right].$$

The fundamental assumption is that features are *conditionally independent* given the class.

Class Prior

The prior probability of class y is estimated from the training data as

$$P(Y = y) = \frac{n_y + \alpha}{n_{\text{tot}} + k\alpha},$$

where n_{tot} is the total number of training instances and n_y is the number of instances belonging to class y.

Discrete Attributes

Let attribute x_i be categorical with possible values $\{c_1, c_2, \ldots, c_m\}$. For each class y, we estimate the conditional probabilities using *Laplace smoothing*. For each feature x_i we have

$$P(c_i \mid Y = y) = \frac{n_{iy} + \alpha}{n_y + m\alpha},$$

where

- n_{iy} = number of training instances in class y with c_i,
- n_y = number of non-missing observations of x_i in class y,
- α = smoothing parameter (typically $\alpha = 1$),
- m = number of distinct categories of attribute i.

If x_i is missing, the term $\log P(x_i \mid Y = y)$ is *skipped*.

Continuous Attributes

If attribute x_i is continuous, it is modeled using a Gaussian distribution:

$$P(x_i \mid Y = y) = \frac{1}{\sqrt{2\pi\sigma_{iy}^2}} \exp\left(-\frac{(x_i - \mu_{iy})^2}{2\sigma_{iy}^2} \right),$$

where

$$\mu_{iy} = \frac{1}{N_{iy}} \sum_{x_i \in D_{iy}} x_i, \quad \sigma_{iy}^2 = \frac{1}{N_{iy}} \sum_{x_i \in D_{iy}} (x_i - \mu_{iy})^2.$$

Here D_{iy} is the set of non-missing values of attribute i in class y. A small variance floor (e.g. $\sigma_{iy}^2 \leftarrow \max(\sigma_{iy}^2, 10^{-6})$) is used to avoid numerical issues.

If x_i is missing, the Gaussian term is skipped.

Prediction

For a new instance $\mathbf{x}$, we compute for each class y:

$$\log P(Y = y) + \sum_{i \in \text{observed}} \log P(x_i \mid Y = y),$$

and predict the class with the highest value.

Dataset	**Accuracy**
Training Set (80%)	66.5%
Testing Set (20%)	63.9%

TABLE II. Performance of the Naive Bayes Algorithm