UiT Norges arktiske universitet

DTE-2501 AI Methods and Applications

Linear methods of classification and regression

Lecture 2/2 – Probabilistic learning model

Tatiana Kravetc
Førsteamanuensis

Office: D2240

Email: tatiana.kravetc@uit.no

Overview

IV Probabilistic model

V Basic terminology

VI Naïve Bayes classifier

VII Example

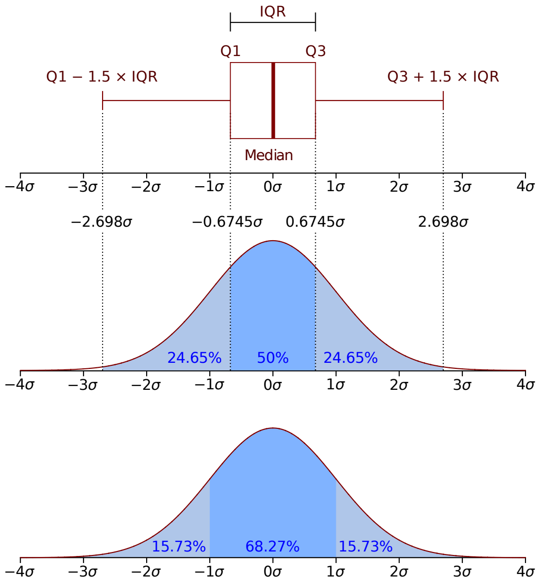
IV Probabilistic model

$x_i \in X^l$ is the available data about a real object.

The data can be:

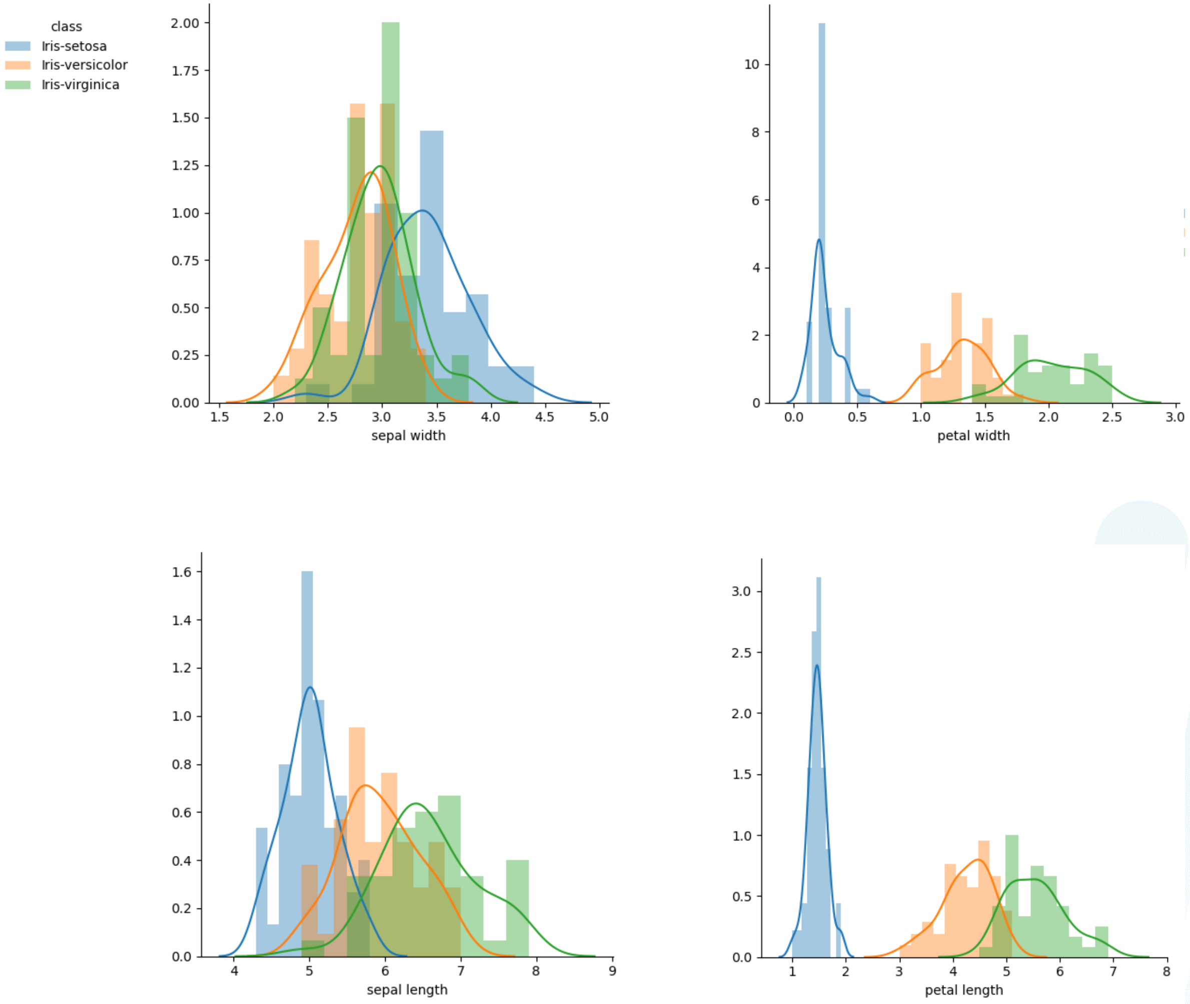
- erroneous
- incomplete

The incorrectness can be eliminated by introducing *a probabilistic formulation* of the problem



$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ is the mean
σ^2 is the variance
σ is the standard deviation



https://en.wikipedia.org/wiki/Probability_density_function

V Basic terminology

- Probability density of $x = P(x)$
- Probability of a specific event $A = P(x = A) = P(A)$
- Probability = (number of desired outcomes) / (total number of possible outcomes)
- Joint probability

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

If two events are *independent*, then $P(A \cap B) = P(A) \cdot P(B)$

- Conditional probability

$$P(A \mid B) = P(A \cap B) / P(B)$$

VI Naïve Bayes classifier

Given classes C_k and a sample $x = (x_1, \dots, x_n)$ having n features to be classified

Bayes' theorem

$$P(C_k|x) = \frac{P(x|C_k)P(C_k)}{P(x)}$$

$P(C_k \mid x)$ is the posterior probability of *class* (C_k, target) given *predictor* (x, features)

$P(C_k)$ is the prior probability of class

$P(x \mid C_k)$ is the likelihood which is probability of predictor given class

$P(x)$ is the prior probability of predictor

Naïve conditional independence assumption

$$P(C_k|x_1, \dots, x_n) = P(x_1|C_k) \cdot P(x_2|C_k) \cdot \dots \cdot P(x_n|C_k) \cdot P(C_k)$$

VII Example. Prediction problem

Historical data

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency table

Weather (x)	Yes (C_0)	No (C_1)
Sunny	3	2
Overcast	4	0
Rainy	2	3
Total	9	5

$P(x)$

=5/14	0.36
=4/14	0.29
=5/14	0.36

$P(C_k)$

=9/14	=5/14
0.64	0.36

Problem: Players will play if weather is sunny. Is this statement correct?

$$P(C_0 \mid \text{sunny}) = \frac{P(\text{ sunny} \mid C_0) \cdot P(C_0)}{P\ (\text{sunny})} = \frac{\frac{3}{9} \cdot 0.64}{0.36} = 0.6$$

https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/