

COMP5318/COMP4318 Machine Learning and Data Mining

Week 4 Tutorial exercises

Naïve Bayes

Exercise 1. Naïve Bayes for data with nominal features (to do in class)

Given is the following dataset where loan default is the class. Predict the class of the following new example using Naïve Bayes:

home owner = no, marital status = married, annual income=very high

	home owner	marital status	income	loan default
1	yes	single	very high	yes
2	no	married	high	yes
3	no	single	medium	no
4	yes	married	very high	no
5	yes	divorced	high	yes
6	no	married	low	no
7	yes	divorced	very high	no
8	no	single	high	yes
9	no	married	medium	no
10	no	single	low	yes

Dataset adapted from, Tan, Steinbach, Karpatne and Kumar, Introduction to Data Mining, Pearson, 2019

Exercise 2. Naïve Bayes for data with numeric features (to do in class)

The same task as in the previous exercise but now annual income is a numeric feature:

	home owner	marital status	income (in K)	loan default
1	yes	single	125	yes
2	no	married	100	yes
3	no	single	70	no
4	yes	married	120	no
5	yes	divorced	95	yes
6	no	married	60	no
7	yes	divorced	220	no
8	no	single	85	yes
9	no	married	75	no
10	no	single	90	yes

Gaussian Naive Bayes assumes that the likelihood $P(x_i|y)$ follows the Gaussian distribution for each x_i given y . Therefore,

$$P(x_i|y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Use Naïve Bayes to predict the class of the following new example:

home owner = no, marital status = married, annual income=120