# Наивный байесовский классификатор

Задача: наблюдаю за своей соседкой по комнате в общежитии. я пытаюсь определить, является ли <<настроение>>, в котором он просыпается, и <<количество часов он спал ночью>> определяет, идет ли он на урок по искусственному интеллекту.

Данные

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| настроение | количество чачов спал | ходить |  |  |  |  |  |  |
| сварливый | три | нет |  |  |  |  |  |  |
| сварливый | три | нет |  |  |  |  |  |  |
| активный | три | нет |  |  |  |  |  |  |
| невеселый | три | да |  |  |  |  |  |  |
| невеселый | восемь | да |  |  | вероятность | | |  |
| невеселый | пять | нет |  |  | настроение | нет | да |  |
| активный | пять | да |  |  | активный | 2 | 2 | 0.29 |
| сварливый | восемь | нет |  |  | сварливый | 3 | 2 | 0.36 |
| сварливый | пять | да |  |  | невеселый | 1 | 4 | 0.36 |
| невеселый | пять | да |  |  | итого | 6 | 8 |  |
| сварливый | пять | да |  |  |  | 0.43 | 0.57 |  |
| активный | пять | да |  |  |  |  |  |  |
| активный | три | нет |  |  |  |  |  |  |
| невеселый | пять | да |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | чистота |  |  |  |  |
|  | настроение | нет | да |  |  |
|  | активный | 2 | 2 |  |  |
|  | сварливый | 3 | 2 |  |  |
|  | невеселый | 1 | 4 |  |  |
|  | итого | 6 | 8 |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  | вероятность | | | | |
|  | настроение | нет | да | p(нет\да) | p(да\нет) |
|  | активный | 2 | 2 | 0.33 | 0.25 |
|  | сварливый | 3 | 2 | 0.5 | 0.25 |
|  | невеселый | 1 | 4 | 0.17 | 0.5 |
|  | итого | 6 | 8 |  |  |

1. **В СЛУЧАЕ ОДНОЙ ФУНКЦИИ**

Теперь предположим, что мы хотим вычислить вероятность того, что мой сосед пойдет на уроки, когда он находится в **“невеселом настроении**”

**вероятность пойти на урок**

*P (да | н) = P (н | да)*P(да) / P (н) .....................(1)

* Вычислить Априорные Вероятности:
  1. **P(невеселый) = 4/14 = 0.286**
  2. **P(да)= 8/14 = 0.57**
* Вычисление Апостериорных Вероятностей:
  1. **P (невеселый |да) = 4/8 = 0.5**
* Априорная и Апостериорная вероятности в уравнении (1)
  1. **P (да | невеселый) = 0.5 \* 0.57 / 0.286 = 0.437**

**Вероятность не пойти на урок**

*P (No |* невеселый*) = P(н | No)*P(No) / P (н) .....................(2)

* Вычислить Априорные Вероятности:

**P(невеселый) = 4/14 = 0.286**

**P(нет)= 6/14 = 0.427**

* Вычисление Апостериорных Вероятностей:

**P(н |нет) = 1/9 = 0.111**

* Ставим Априорная и Апостериорная вероятности в уравнении (2)

**P (No | невеселый) = 0.111\* 0.427/0.286=0.166**

**ДА(0.437)>НЕТ(0.166) мой сосед пойдет на урок по СИИ даже если у него невеселые настроение**

1. **В СЛУЧАЕ НЕСКОЛЬКИХ ФУНКЦИЙ**

теперь давайте подсчитаем вероятность того, что мой сосед пойдет на уроки, если он будет в <<**невеселый настроении**>> и <<**проспит всего 5 часов>>**

**ПОЙДЕТ НА УРОКИ (<<невеселый настроении**>> и <<**проспит всего 5 часов>>**

А.**P(пойти=да|настроение=невеселый,кол.ч.спал=пять)= P(настроение=невеселый, кол.ч.спал=пять| пойти= да)P(пойти=да)**

В**.P(настроение=невеселый,кол.ч.спал=пять|пойти=да)= P(невеселый |да) P(пять |да)**

Вычислить Априорные Вероятности:

**P(Yes)= 8/14 = 0.57**

Вычисление Апостериорных Вероятностей:

**P(невеселый |да) = 4/8 = 0.5**

**P(пять |да) = 5/8 = 0.625**

Ставим Априорная и Апостериорная вероятности в уравнении (**В**)

**P(настроение=невеселый, кол.ч.спал=пять | пойти= да) = 0.5 \* 0.625 = 0.156**

ставим Априорная и Апостериорная вероятности в уравнении (**А**)

**P(пойти= да | настроение=невеселый, кол.ч.спал=пять) = 0.156\*0.57 = 0.089**

**НЕ ПОЙДЕТ НА УРОК (невеселые настроении**>> и <<**проспит всего 5 часов>>**

**C.P(пойти=нет|настроение=невеселый,кол.ч.спал=пять)= P(настроение=невеселый,кол.ч.спал=пять|пойти=да)P(пойти=нет)**

**D.P(настроение=невеселый,кол.ч.спал=пять|пойти=нет)= P(настроение=невеселый |пойти=нет) P(кол.ч.спал=пять | пойти=нет)**

Вычислить Априорные Вероятности: **P(нет)= 6/14 = 0.429**

Вычисление Апостериорных Вероятностей: P(настроение=**невеселый |пойти=нет) = 1/6 = 0.167 P(кол.ч.спал=пять | пойти=нет)=1/6=0.167**

Ставим Априорные Вероятности в уравенение (**D**) **P(настроение=невеселый, кол.ч.спал=пять |пойти= нет) = 0.67\* 0.167= 0.112**

ставим Априорная и Апостериорная вероятности в уравнении (**C**) **P (пойти=нет|настроение=невеселый,кол.ч.спал=пять|)= 0.112\*0.429=0.048**

вероятность “да”> вероятность нет,можно сказать что мой сосед пойдет на уроки если у него веселый настроение и если он спал

**Достатки**

* Это не только простой подход, но и быстрый и точный метод прогнозирования.
* Он может эффективно работать с большим набором данных.
* Может использоваться с несколькими задачами прогнозирования классов.

**Недостатки**

* Не заметил их

**Код в python**

#В этом примере я буду использовать набор данных (который я собрал) с тремя столбцами:

#(настроение), (количество сна),(пойти или не пойти).

#Первые два-это особенности (настроение, часы сна) , а второй-ярлык " пойти"

#сварлтвый(grumpy),активный(enegetic),невеселый(sad)

mood=['grumpy','grumpy','energetic','sad','sad','sad','energetic','grumpy','grumpy',

'sad','grumpy','energetic','energetic','sad']

hours=['three','three','three','five','eight','eight','eight','five','eight','five','five','five','three','five']

go=['No','No','Yes','Yes','Yes','No','Yes','No','Yes','Yes','Yes','Yes','Yes','No']

# Во-первых, вам нужно преобразовать эти строковые метки в числа. например: "сварлтвый", "активный", "невеселый"

# как 0, 1, 2. Это называется кодированием меток. Scikit-learn предоставляет библиотеку LabelEncoder

# для кодирования меток со значением от 0 до одного меньше числа дискретных классов

from sklearn import preprocessing

#creating labelEncoder

le = preprocessing.LabelEncoder()

# Преобразование строковых меток в числа.

mood\_encoded=le.fit\_transform(mood)

print (mood\_encoded)

# Аналогично, вы также можете кодировать столбцы <<часы сна>> и <<пойти>>.

hours\_encoded=le.fit\_transform(hours)

label=le.fit\_transform(go)

print (hours)

print (go)

#Теперь объедините обе функции (настроение и количество часов) в одну переменную (список кортежей).

features=[(2, 1), (2, 1), (0, 1), (1, 2), (1, 0), (1, 0), (0, 0), (2, 2), (2, 0), (1, 2), (2, 2), (0, 2), (0, 1), (1, 2)]

print (features)

#Import Гауссовой модели упрощенного алгоритма Байеса

from sklearn.naive\_bayes import GaussianNB

#Создание Гауссова классификатора

model = GaussianNB()

# Тренируйте модель с помощью обучающих наборов

model.fit(features,label)

#Прогнозируемый Выход

predicted= model.predict([[0,2]]) # 0:невеселый,sad, 2:пять часов сна 5

print ('predicted :' , predicted)

**output**

[1 1 0 2 2 2 0 1 1 2 1 0 0 2]

['three', 'three', 'three', 'five', 'eight', 'eight', 'eight', 'five', 'eight', 'five', 'five', 'five', 'three', 'five']

['No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'Yes', 'No']

[(2, 1), (2, 1), (0, 1), (1, 2), (1, 0), (1, 0), (0, 0), (2, 2), (2, 0), (1, 2), (2, 2), (0, 2), (0, 1), (1, 2)]

predicted : [1]

**[1] означает что мой сосед пойдет на уроки**