Plan of Attack

22 June 2023 13:59

Naive Bayes → 96

Naive Bayes → classification → Textual data

✓

Baseline model

→ Xgboost

Deep → 96.5%

spam filter

sentiment analysis

1

✓ Basics
✓ code
✓ case study

2

Advanced topics

Intuition

22 June 2023 13:59

Play Tennis Toy

binary classification

Naive $\boxed{\text{Bayes}}$ theorem

D|
14 days

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No ✗
Sunny	Hot	High	True	No ✗
Overcast	Hot	High	False	Yes ✓
Rainy	Mild	High	False	Yes ✓
Rainy	Cool	Normal	False	Yes ✓
Rainy	Cool	Normal	True	No ✗
Overcast	Cool	Normal	True	Yes ✓
Sunny	Mild	High	False	No ✗
Sunny	Cool	Normal	False	Yes ✓
Rainy	Mild	Normal	False	Yes ✓
Sunny	Mild	Normal	True	Yes ✓
Overcast	Mild	High	True	Yes ✓
Overcast	Hot	Normal	False	Yes ✓
Rainy	Mild	High	True	No ✗

{Sunny, Cool, Normal, True} = W

query point

Y/N

2 prob

Yes 0.82 ✓

No 0.18

Bayes theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$$\boxed{P(Yes|W)} = \frac{\boxed{P(W|Yes)}\,P(Yes)}{0}$$

4

9/14

$$P(No|W) = \frac{P(W|NO)\,P(NO)}{0}$$

5/14

$\boxed{P(\text{sunny} \cap \text{cool} \cap \text{Normal} \cap \text{true} \mid \text{Yes})}$

$\boxed{P(W|Yes)}$ $\quad P(W|NO)$

Bayes theorem

Naive - ganna

$P(\text{sunny} \cap \text{cool} \cap \text{Normal} \cap \text{True} \mid NO) = 0$

$\boxed{P(\text{sunny}|\text{Yes})}\ P(\text{cool}|\text{yes})\ P(\text{Normal}|\text{Yes})\ P(\text{True}|\text{yes})$

$2/9 \times 3/9 \times 6/9 \times 3/9 \times 9/14$

$P(\text{Sunny}|NO)\ P(\text{Cool}|NO)\ P(\text{Normal}|NO)\ P(\text{True}|NO) \times \frac{5}{14}$

Mathematical Formulation

22 June 2023 13:59

y N

$x_1 \; x_2 \cdots x_n \quad y$ multiclass classification $\underbrace{1, 2, 3 \cdots k}$ k classes

$\rightarrow \langle x_1, x_2, x_3 \cdots x_n \rangle \rightarrow$ naive bayes k prob

$$P(y_1 \mid X_T) = P(X_T \mid y_1)\, P(y_1)$$

$$P(y_2 \mid X_T) \rightarrow P(X_T \mid y_2)\, P(y_2)$$

$$\vdots$$

$$\boxed{P(y_k \mid X_T)} \rightarrow P(X_T \mid y_k)\, P(y_k)$$

$$X_T = \langle x_1, x_2 \cdots x_n \rangle$$

$\cap$,

$$P(y_k \mid X_T) = P(X_T \mid y_k)\, P(y_k)$$

$$= P(x_1 \cap x_2 \cap x_3 \cap \cdots \cap x_n \mid y_k)\, P(y_k)$$

$$= P(\overset{A}{\boxed{x_1, x_2, x_3 \cdots, x_n}} \mid \overset{B}{\boxed{y_k}})\, P(y_k)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

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

$$= \frac{P(x_1, x_2, x_3 \cdots x_n, y_k)}{\cancel{P(y_k)}} \cancel{P(y_k)}$$

$$= P(\overset{A}{\boxed{x_1}} \overset{\cap}{,} \overset{B}{\boxed{x_2, x_3, \cdots x_n, y_k}})$$

$$= P(x_1 \mid x_2, x_3, \cdots x_n, y_k)\, P(\overset{A}{\boxed{x_2}}, \overset{B}{\boxed{x_3, \cdots, x_n, y_k}})$$

$$P(x_2 \mid x_3, x_4, \cdots x_n, y_k)\, P(\overset{A}{\boxed{x_3}}, \overset{B}{\boxed{x_4 \cdots x_n, y_k}})$$

$$= P(x_1 \mid x_2, x_3 \cdots x_n, y_k)\, P(x_2 \mid x_3, x_4 \cdots y_k) \cdots P(x_{n-1} \mid x_n, y_k)\, P(x_n \mid y_k)\, P(y_k)$$

Naive assumption $\rightarrow$ features are independent of each other

x_1	x_2	x_3	$\cdots$	x_n

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

$$P(A \mid B \cap C) = P(A \mid B, C) = P(A \mid C)$$

X_1	X_2	X_3	...	X_n

$$P(A \mid B \cap C) = P(A \mid B, C) = P(A \mid C)$$

$$y_1 = \boxed{P(X_1 \mid y_k)}\, P(X_2 \mid y_k)\, P(X_3 \mid y_k) \ldots \ldots P(X_{n-1} \mid y_k)\, P(X_n \mid y_k)\, P(y_k)$$

y_2

$\vdots$

y_k

max $\rightarrow$ maximum a posteriori Rule

(MAP)

Code

22 June 2023 14:00

No, Yes

2

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Sunny, Hot, High, False

Training → Probabilities → Testing

dictionary

Outlook → Sunny | Overcast | Rainy

$\boxed{3 \times 2} = 6$ prob 3×2 2×2

P(Sunny | Yes)
P(sunny | NO)
P(overcast | Yes)
P(overcast | No)
P(Rainy | Yes)
P(Rainy | No)

P(Hot | Y)
P(Hot | N)
P(Mild | Y)
P(Mild | N)
P(Cool | Y)
P(Cool | N)

P(H | Y)
P(H | N)
P(N | Y)
P(N | No)

How Naïve Bayes handles numerical data?

22 June 2023 16:51

age

27

61

52

⋮

married

σ, μ

Y →

N → σ, μ

Y

{ 55 } → Y, N?

68 , 72

category

$P(Y \mid 55) = P(55 \mid Y)\, P(Y)$ → 0.62

0 0

$P(N \mid 55) = P(55 \mid N)\, P(N)$

gaussian dist

$\mu\ \sigma$

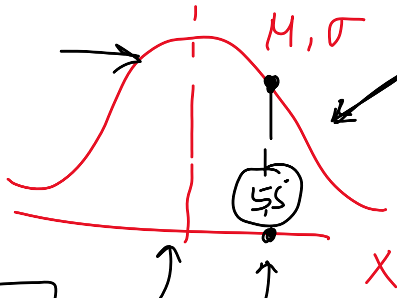
55

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

Probabaq

prob

0.62



What is data is not Gaussian?

22 June 2023 16:52

1. Data Transformation: Depending on the nature of your data, you could apply a transformation to make it more normally distributed. Common transformations include the logarithm, square root, and reciprocal transformations.

10-20 | 20-30 | -

2. Alternative Distributions: If you know or suspect that your data follow a specific non-normal distribution (e.g., exponential, Poisson, etc.), you can modify the Naïve Bayes algorithm to assume that specific distribution when calculating the likelihoods.

→ Tryout

3. Discretization: You can turn your continuous data into categorical data by binning the values. There are various ways to decide on the bins, including equal width bins, equal frequency bins, or using a more sophisticated method like k-means clustering. Once your data is binned, you can use the standard Multinomial or Bernoulli Naïve Bayes methods.

binning

4. Kernel Density Estimation: A non-parametric way to estimate the probability density function of a random variable. Kernel density estimation can be used when the distribution is unknown.

5. Use other models: If none of the above options work well, it may be best to consider a different classification algorithm that doesn't make strong assumptions about the distributions of the features, such as Decision Trees, Random Forests, or Support Vector Machines

Naïve Bayes on Text Data

22 June 2023 14:13

Sentiment Analysis

transform

review	sentiment
	+
	—
	+

5000

→ 1) Text Preprocessing

→ 2) Vectorize → text → numbers
- BOW →
- OHE
- TFIDF
- Embeddings

numbers

→ 3) apply naive bayes

BOW

10L words → 35000 uniques

↳ 5000 most frequently used words

reviews	sentiment
→ I liked the movie	+ve
really hated the movie	—ve
—	
—	
—	

(50000, 5000)

↳ shaped

hate	like	movie	actor	. . . —
0	1	1	0	. . .
2	0	3	2	. . . . .
3	2	1	1	
⋮				

Adj - - great movie → numbers

(0 2 1 . . . . 5) → (1,5000)

vector

+ve —ve

$$\boxed{P(+ve \mid hate=0,\ like=2,\ \ldots)} \longrightarrow \underline{0.37}$$

$$P(-ve \mid hate=0,\ like=2,\ \ldots) \longrightarrow \boxed{0.23}$$

$\rightarrow$ +ve positive ~~sentiment~~

$\boxed{P(}hate=0 \mid +ve)\ P(like=2 \mid +ve) \ldots$

$\hookrightarrow p$