CIS 490 Machine Learning

Lecture 12

Instructor: (Julia) Hua Fang

1

2

Last time

- R: Running CART in R studio.
- Introduction to Exam I and Project II Classification.

2

3

Reminder:

- March 21: Sectional project II code, report and slide submission
- March 22: Sectional Project II Zoom presentation

3

Outline

- Supervised Learning

- **Classification**
  - Bayesian Classifiers: Naïve Bayes
    - -- Bayesian vs. Frequentist
    - -- Review Bayesian Theorem
    - -- Naïve Bayes classification: two examples: a. X attributes are discrete; b X attributes have continuous variables.
  - Run Naïve Bayes in R Studio

4

Frequentist vs. Bayesian: Facts

Frequentist:

- Data are a repeatable random sample
  - there is a frequency
- Underlying parameters remain constant during this repeatable process
- Parameters are fixed

Bayesian:

- Data are observed from the realized sample.
- Parameters are unknown and described probabilistically
- Data are fixed

5

Frequentist vs. Bayesian: Inference

Frequentist:

- Point estimates and standard errors or 95% confidence intervals.
- Deduction from P (data | H0), by setting ***α*** in advance.
  - Accept H1 if P (data | H0) < ***α***.
  - Accept H0 if P (data | H0) ≥ ***α***.

Note: H0 is called null hypothesis; H1 is called alternative hypothesis

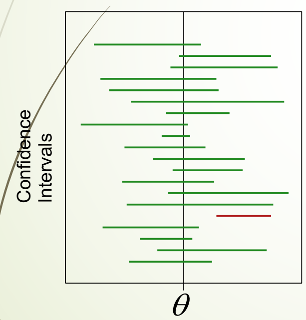
Bayesian:

- Induction from P (θ | data), starting with P (θ).
- Broad descriptions of the posterior distribution such as means and quantiles.
- Highest posterior density intervals indicating region of highest posterior probability.

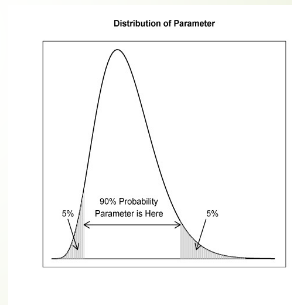
6

Frequentist vs. Bayesian: Interpretations of Confidence

Frequentist: A collection of intervals with 90% of them containing the true parameter



Bayesian: An interval that has a 90% chance of containing the true parameter.



7

Frequentist vs. Bayesian: Summary

- Frequentist: P (data | H0) is the sampling distribution of the data given the parameter
- Bayesian: P (θ) is the **prior** distribution of the parameter (before the data are seen); P (θ | data) is the **posterior** distribution of the parameter

Note: H0 is called null hypothesis

8

Where Do Priors Come From?

9

Where Do Priors Come From?

- Previous studies, published work
- Researcher intuition
- Substantive Experts
- Convenience
- Other data sources

10

Bayesian classification: Review Bayes Theorem

- Conditional probability: C= Class label; **x** = features/attributes.

$$p(C \mid \mathbf{x}) = \frac{p(\mathbf{x}, C)}{p(\mathbf{x})} \qquad p(\mathbf{x} \mid C) = \frac{p(\mathbf{x}, C)}{p(C)}$$

- Bayes theorem:

$$\underbrace{p(C \mid \mathbf{x})}_{\text{posterior probability}} = \frac{\overbrace{p(\mathbf{x} \mid C)}^{\text{likelihood}}\,\overbrace{p(C)}^{\text{prior probability}}}{\underbrace{p(\mathbf{x})}_{\text{evidence}}}$$

11

Example of Bayes theorem: in-class exercise

- Given:
  - A doctor knows that meningitis causes stiff neck 50% of the time (hint: Likelihood)
  - Prior probability of any patient having meningitis is 1/50,000
  - Prior probability of any patient having stiff neck is 1/20 (hint: Evidence)

Question: If a patient has stiff neck, what's the probability he/she has meningitis? Let M= Meningitis and S= Stiff neck, then

$$p(M \mid S) = \frac{p(S \mid M)\,p(M)}{p(S)} = ?$$

12

Answers:

$$p(M \mid S) = \frac{p(S \mid M)\,p(M)}{p(S)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002$$

13

Bayesian classification: in general

$$\boxed{p(C_i \mid x_1, x_2, \ldots, x_n) = \frac{p(x_1, x_2, \ldots, x_n \mid C_i)\,p(C_i)}{p(x_1, x_2, \ldots, x_n)}}$$

$$p(C_i \mid x_1, x_2, \ldots, x_n) \propto p(x_1, x_2, \ldots, x_n \mid C_i)\,p(C_i)$$

Notation: $\propto$ means proportional to, up to a constant factor

14

Special case: Naïve Bayes classifier

$$p(C_i \mid x_1, x_2, \ldots, x_n) \propto p(x_1, x_2, \ldots, x_n \mid C_i)\, p(C_i)$$

Naïve Bayes assume **independence** among attributes x_j when class is given:

$$p(x_1, x_2, \ldots, x_n \mid C_i) = p(x_1 \mid C_i)\, p(x_2 \mid C_i) \ldots p(x_n \mid C_i)$$

15

Example of Naïve Bayes classifier: in class exercise

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

_X_: attributes

Two classes:

_M_: class = mammal

_N_: class = non-mammal

New case:

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P (C | X) =?

16

Example of Naïve Bayes classifier: How to

***X*: attributes**

***M*: class = mammal**

***N*: class = non-mammal**

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Given the features of this new case and data in the table, this case is mammal or not. Mathematically,

P (Class = Mammal | X)

or

P(Class = non-Mammal | X)

New case:

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P (Class = Mammal | X) ∝ *p*(x | Class = mammal) *p*(Class = mammal)

Vs.

P(Class = non-Mammal | X) ∝ *p*(x | Class = non-mammal) *p*(Class = non-mammal)

17

Example of Naïve Bayes classifier: How to

Reminder: Naïve Bayes assume independence among attributes x_j when class is given

New case:

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

***X*: attributes**

***M*: class = mammal**

***N*: class = non-mammal**

- P (Class = Mammal | X) ∝ *p*(x | Class = mammal) *p*(Class = mammal)

p(x | Class = mammal) = *p*(GiveBirth = Yes | Class = mammal) × *p*(CanFly=No | Class = mammal) × *p*(LiveinWater=Yes | Class = mammal) × *p*(HaveLegs = No | Class = mammal)

p(Class = mammal) = N_*mammal*/*N*

vs.

- P(Class = non-Mammal | X) ∝ *p*(x | Class = non-mammal) *p*(Class = non-mammal)

p(x | Class =non-mammal) = *p*(GiveBirth = Yes | Class = non-mammal) × *p*(CanFly=No | Class = non-mammal) × *p*(LiveinWater=Yes | Class = non-mammal) × *p*(HaveLegs = No | Class = non-mammal)

p(Class = non-mammal) = N_*non-mammal*/*N*

18

Example of Naïve Bayes classifier: How to

New case:

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

***X*: attributes**

***M*: class = mammal**

***N*: class = non-mammal**

$$p(X \mid M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$p(X \mid N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$p(X \mid M)p(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$p(X \mid N)p(N) = 0.0042 \times \frac{13}{20} = 0.0027$$

- P (Class = Mammal | X) $\propto$ *p*(x | Class = mammal) *p*(Class = mammal) and, *p*(*X* | *M*) *p*(*M*) = 0.021
- P(Class = non-Mammal | X) $\propto$ *p*(x | Class = non-mammal) *p*(Class = non-mammal) and, *p*(*X* | *N*) *p*(*N*) = 0.0027

Since P (Class = Mammal|X) > P(Class = non-Mammal|X), this new case is classified as mammal

19

Naïve Bayes: continuous attributes (II)

For continuous attributes:

- Discretize the range into bins
  - What type of attribute will be? (ordinal or binary)
- Two-way split: ($x_i < v$) or ($x_i > v$)
  - What type of attribute will be? (ordinal or binary)
- Probability density estimation:
  - assume attribute follows some standard parametric probability distribution (usually a Gaussian)
  - use data to estimate parameters of distribution (e.g. mean and variance)

20

In-class exercise: Bayes with continuous variable

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- For (Income | Class = No):
  - sample mean = 110k
  - sample variance = 2975

Let's assume income follows Gaussian

21

In-class exercise: Bayes with continuous variable

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Gaussian distribution:

$$P(x_j \mid C_i) = \frac{1}{\sqrt{2\pi\sigma_{ji}^2}} e^{-\frac{(x_j-\mu_{ji})^2}{2\sigma_{ji}^2}}$$

  - one for each (x_j, C_i) pair

22

Recall: Naïve Bayes classifier

$$p(C_i \mid x_1, x_2, \ldots, x_n) \propto p(x_1, x_2, \ldots, x_n \mid C_i)\, p(C_i)$$

Naïve Bayes assume independence among attributes x_j when class is given:

$$p(x_1, x_2, \ldots, x_n \mid C_i) = p(x_1 \mid C_i)\, p(x_2 \mid C_i) \ldots p(x_n \mid C_i)$$

23

In-class exercise: Bayes with continuous variable

Given a Test/new Record:

$$\mathbf{x} = (\text{Refund} = \text{No}, \text{Status} = \text{Married}, \text{Income} = 120\text{K})$$

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

naive Bayes classifier:

- (Refund = Yes | No) = 3/7
- (Refund = No | No) = 4/7
- (Refund = Yes | Yes) = 0
- (Refund = No | Yes) = 1
- (Marital Status = Single | No) = 2/7
- (Marital Status = Divorced | No) = 1/7
- (Marital Status = Married | No) = 4/7
- (Marital Status = Single | Yes) = 2/7
- (Marital Status = Divorced | Yes) = 1/7
- (Marital Status = Married | Yes) = 0

For Taxable Income:
If Class = No: sample mean = 110
sample variance = 2975
If Class = Yes: sample mean = 90
sample variance = 25

24

In-class exercise: Bayes with continuous variable

Given a Test/new Record:

$$\mathbf{x} = (\text{Refund} = \text{No}, \text{Status} = \text{Married}, \text{Income} = 120\text{K})$$

naive Bayes classifier:

- (Refund = Yes | No) = 3/7
- (Refund = No | No) = 4/7
- (Refund = Yes | Yes) = 0
- (Refund = No | Yes) = 1
- (Marital Status = Single | No) = 2/7
- (Marital Status = Divorced | No) = 1/7
- (Marital Status = Married | No) = 4/7
- (Marital Status = Single | Yes) = 2/7
- (Marital Status = Divorced | Yes) = 1/7
- (Marital Status = Married | Yes) = 0

For Taxable Income:
If Class = No: sample mean = 110
sample variance = 2975
If Class = Yes: sample mean = 90
sample variance = 25

- *p*(**x** | Class = No) = *p*(Refund = No | Class = No)
  × *p*(Married | Class = No)
  × *p*(Income = 120K | Class = No)
  = 4/7 × 4/7 × 0.0072 = 0.0024

  p(No) =7/10

- *p*(**x** | Class = Yes) = *p*(Refund = No | Class = Yes)
  × *p*(Married | Class = Yes)
  × *p*(Income = 120K | Class = Yes)
  = 1 × 0 × (1.2 × 10^{-9})= 0

  p(Yes) = 3/10

***p*(x | No) *p*(No) > *p*(x | Yes) *p*(Yes)**

therefore *p*(No | x) > *p*(Yes | x), this new is classified No Evade.

See next slide for how to compute p(Income = 120K | Class = No/Yes)

25

In-class exercise: Bayes with continuous variable

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Given (Income | Class = No) is Gaussian and

- sample mean = 110k
- sample variance = 2975

$$P(x_j \mid C_i) = \frac{1}{\sqrt{2\pi\sigma_{ji}^2}} e^{-\frac{(x_j-\mu_{ji})^2}{2\sigma_{ji}^2}}$$

p(Income = 120K | Class = No) =?

$$p(\text{ Income} = 120 \mid \text{No }) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120k-110k)^2}{2(2975)}} = 0.0072$$

$$\sqrt{2975} = 54.54$$

Likewise, p(Income = 120K | Class = Yes) = 1.2 × 10^{-9}

26

Suppl.: Naïve Bayes classifier

- Problem: if one of the conditional probabilities is zero, then the entire expression becomes zero.
- This is a significant practical problem, especially when training samples are limited.
- Ways to improve probability estimation:

$$\text{Original}: p(x_j \mid C_i) = \frac{N_{ji}}{N_i}$$

$$\text{Laplace}: p(x_j \mid C_i) = \frac{N_{ji}+1}{N_i+c}$$

$$\text{m-estimate}: p(x_j \mid C_i) = \frac{N_{ji}+mp}{N_i+m}$$

27

Summary of naïve Bayes

- Robust to isolated noise samples.
- Handles missing values by ignoring the sample during probability estimate calculations.
- Robust to irrelevant attributes.
- *NOT* robust to redundant attributes.
  - Independence assumption does not hold in this case.

    Remedy: Use other techniques such as Bayesian Belief Networks (BBN).

28

29

R: Running Naïve Bayes in R studio.

R: Run Naïve Bayes in Rstudio using IRIS data

Let's go through the instruction file "R_NaiveBayes.docx" posted with LS12 slides at myCourses.

29

30

Learning Activity 3 (LA3): 30 points; Due: Mar 28th.

- Review LS12 Slides and follow the instruction file "R_NaiveBayes.docx", implement the following:
- set.seed(490) and split the training and testing set, as 50% (training) vs. 50% (testing)
- Run naiveBayes() on training, testing, and full data set, respectively:
  - ✓ Output confusion table for each dataset and compute sensitivity, specificity, ppv and npv for each dataset
  - ✓ Create ROC curves, including micro- and macro-average curves using **testing** data set
- Refer to "**Results Table**" in "R_NaiveBayes.docx", create the comparison table summarizing each evaluation metric across the testing, training, and full datasets.

Note: The late policy does NOT apply to Learning Activity (LA) Assignments. LAs are not group assignment. To receive your score, each individual must submit your Complete work on Time.

30

Learning Activity 4(LA4): 30 points, Due: April 4th.

31.

Read the two attached files at myCourses: one PPT file called "CIS490_Reading_TasteofDNN.pdf" and one called "Chexnet_Ng.pdf", then answer the following questions:

1. Describe the major concept difference between traditional neural network and deep learning.
2. What are typical applications of deep learning?
3. What make Deep learning state of the Art?
4. What are the three technology enablers make this degree of accuracy possible for deep learning?
5. Describe Deep neural network structure.
6. Describe coevolutionary neural network (CNN) structure, the three types of operations of its Feature Detection Layers, and classification layers.
7. What is AlexNet?
8. How many layers does CheXNet have? How many images are used for training, validation and test, respectively? What are main contributions of CheXNet?

Note: Don't copy slides; use your own language. The late policy does NOT apply to Learning Activity (LA) Assignments. LAs are not group assignment. To receive your score, each individual must submit your Complete work on Time.

31