SIGAI

Special Interest Group: Artificial Intelligence

27 January 2018

Introductions!

- Name
- Education Year & Major
- Number of years programming
- Favorite Show/Movie

Add Yourself to the Mailing List

If someone want to subscribe to the mailing list, send an email to "**sigai-request@freelists.org**" with "**subscribe**" in the subject.

Get added to the Facebook group chat

Add me or Ray on Facebook and let us know to add you to the chat.

Vote for Meeting Times!

Poll: https://goo.gl/MnZ3kg

As a CS student, you can get a keyfob from the Lawson mailroom which will give you after-hours/weekend access to the Lawson building, as well as the student activities center in HAAS G072.

Getting a new key fob for access to Lawson after-hours

To get a keyfob, simply email **cs-workstudy@science.purdue.edu** . Here's a template that might be helpful:

> Subject: Lawson Keyfob
>
> To whom it may concern,
>
> I am a computer science student enrolled in [CS courses here]. I would like a key fob to access the Lawson building after-hours.
>
> Thank you, [Your name]

Machine Learning

Definition:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. **The primary aim is to allow the computers learn automatically** without human intervention or assistance and adjust actions accordingly.

Machine Learning

Machine learning algorithms are often categorized as supervised or unsupervised.

- **Supervised machine learning algorithms** rely on each example in data to be labeled or associated with a class. These algorithms use the analysis of a known training dataset, to create an inferred function to make predictions about the output values of new examples.

- In contrast, **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labeled. These algorithms are designed to infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Naïve Bayes

Naïve Bayes

- Bayes classification

$$P(c \mid \mathbf{x}) \propto P(\mathbf{x} \mid c)P(c) = P(x_1, \cdots, x_n \mid c)P(c) \ \text{ for } c = c_1, \ldots, c_L.$$

Difficulty: learning the joint probability $P(x_1, \cdots, x_n \mid c)$ is infeasible!

Naïve Bayes

- Bayes classification

$$P(c\,|\,\mathbf{x}) \propto P(\mathbf{x}\,|\,c)P(c) = P(x_1,\cdots,x_n\,|\,c)P(c) \;\; \text{for } c = c_1,\ldots,c_L.$$

Difficulty: learning the joint probability $P(x_1,\cdots,x_n\,|\,c)$ is infeasible!

- Naïve Bayes classification
  - Assume all input features are class conditionally independent!

$$\begin{aligned} P(x_1,x_2,\cdots,x_n\,|\,c) &= \underline{P(x_1\,|\,x_2,\cdots,x_n,c)}P(x_2,\cdots,x_n\,|\,c) \\ &= \underline{P(x_1\,|\,c)}P(x_2,\cdots,x_n\,|\,c) \\ &= P(x_1\,|\,c)P(x_2\,|\,c)\cdots P(x_n\,|\,c) \end{aligned}$$

Applying the independence assumption

Naïve Bayes

- Bayes classification

$$P(c\,|\,\mathbf{x}) \propto P(\mathbf{x}\,|\,c)P(c) = P(x_1,\cdots,x_n\,|\,c)P(c) \quad \text{for } c = c_1,\ldots,c_L.$$

Difficulty: learning the joint probability $P(x_1,\cdots,x_n\,|\,c)$ is infeasible!

- Naïve Bayes classification
  - Assume all input features are class conditionally independent!

$$P(x_1,x_2,\cdots,x_n\,|\,c) = \underline{P(x_1\,|\,x_2,\cdots,x_n,c)}P(x_2,\cdots,x_n\,|\,c)$$
$$= \underline{P(x_1\,|\,c)}P(x_2,\cdots,x_n\,|\,c)$$
$$= P(x_1\,|\,c)P(x_2\,|\,c)\cdots P(x_n\,|\,c)$$

Applying the independence assumption

  - Apply the MAP classification rule: assign $\mathbf{x}' = (a_1, a_2, \cdots, a_n)$ to c^* if

$$\underbrace{[P(a_1\,|\,c^*)\cdots P(a_n\,|\,c^*)]}_{\text{estimate of } P(a_1,\cdots,a_n\,|\,c^*)}P(c^*) > \underbrace{[P(a_1\,|\,c)\cdots P(a_n\,|\,c)]}_{\text{esitmate of } P(a_1,\cdots,a_n\,|\,c)}P(c), \quad c \neq c^*, c = c_1,\cdots,c_L$$

Example

- Example: Play Tennis

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example

- Learning Phase

Outlook	Play=*Yes*	Play=*No*
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	Play=*Yes*	Play=*No*
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

Humidity	Play=*Yes*	Play=*No*
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=*Yes*	Play=*No*
Strong	3/9	3/5
Weak	6/9	2/5

$P(\text{Play}=Yes) = 9/14$ $P(\text{Play}=No) = 5/14$

Example

- Test Phase
  - Given a new instance, predict its label

    x'=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)

Example

- Test Phase
  - Given a new instance, predict its label

    x′=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)
  - Look up tables achieved in the learning phrase

P(Outlook=*Sunny*|Play=*Yes*) = 2/9
P(Temperature=*Cool*|Play=*Yes*) = 3/9
P(Huminity=*High*|Play=*Yes*) = 3/9
P(Wind=*Strong*|Play=*Yes*) = 3/9
P(Play=*Yes*) = 9/14

P(Outlook=*Sunny*|Play=*No*) = 3/5
P(Temperature=*Cool*|Play==*No*) = 1/5
P(Huminity=*High*|Play=*No*) = 4/5
P(Wind=*Strong*|Play=*No*) = 3/5
P(Play=*No*) = 5/14

Example

- Test Phase
  - Given a new instance, predict its label

    x′=(Outlook=*Sunny*, Temperature=*Cool*, Humidity=*High*, Wind=*Strong*)
  - Look up tables achieved in the learning phrase

P(Outlook=*Sunny*|Play=*Yes*) = 2/9	P(Outlook=*Sunny*|Play=*No*) = 3/5
P(Temperature=*Cool*|Play=*Yes*) = 3/9	P(Temperature=*Cool*|Play==*No*) = 1/5
P(Huminity=*High*|Play=*Yes*) = 3/9	P(Huminity=*High*|Play=*No*) = 4/5
P(Wind=*Strong*|Play=*Yes*) = 3/9	P(Wind=*Strong*|Play=*No*) = 3/5
P(Play=*Yes*) = 9/14	P(Play=*No*) = 5/14

  - Decision making with the MAP rule

$$P(Yes|\mathbf{x}') \approx [P(Sunny|\mathrm{Yes})P(Cool|\mathrm{Yes})P(High|\mathrm{Yes})P(Strong|\mathrm{Yes})]P(\mathrm{Play}=Yes) = 0.0053$$

$$P(No|\mathbf{x}') \approx [P(Sunny|\mathrm{No})\ P(Cool|\mathrm{No})P(High|No)P(Strong|No)]P(\mathrm{Play}=No) = 0.0206$$

Given the fact $P(Yes|\mathbf{x}') < P(No|\mathbf{x}')$, we label $\mathbf{x}'$ to be "*No*".

Naive Bayes : Pros and Cons

Pros

- It is easy and fast to predict class of test data set. It also perform well in multi class prediction
- It perform well in case of categorical input variables compared to numerical variable(s).

Cons

- Zero conditional probability, i.e if no example contains the feature value.
  - Fix: Laplace Smoothing
- Conditional Probability independence assumption
  - This is why Naive Bayes is known to be a **bad estimator**

References

- http://www.expertsystem.com/machine-learning-definition/
- http://syllabus.cs.manchester.ac.uk/ugt/2017/COMP24111/materials/slides/Naive-Bayes.pdf
- https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/