Naïve Bayes Classifiers

Dae-Ki Kang

Probability

- What is the probability that you get hit in the head with bird droppings when you go outside?
  - Initially shown in one episode of Korean TV show
  - Asked in one BBS of an elite group in Korea
    - no one gave the right answer
- It is difficult to define “probability”
  - As difficult as defining “science”, “history”, etc.
- 16 different definitions
- Frequentist vs. Bayesian

Bayesian Probability

- Can assign probability on any statement
- Can be used to show the degree of one's belief, or degree of one's knowledge.
- Example of subjectiveness: Alice, Bob, and Carol
  - Alice tosses a new coin 10 times → 3/10 heads
  - Bob tosses the coin 10 times → 4/10 heads
  - Carol has watched them all → 7/10 heads
- Probabilities in Monty Hall
  - MC knows where the car is → 1
  - Guest asked again after MC opened the door → 2/3
  - Viewer started watching the show after MC opened the door → 1/2

Bayes Theorem

- Probability – P(A)
- Conditional Probability – P(A|C)
- P(A|B) = P(A,B)/P(B)
- P(B|A) = P(A,B)/P(A)
- P(A|B) = {P(B|A)P(A)}/P(B)
- P(C|D) = {P(D|C)P(C)}/P(D)
  - P(C|D) : The probability of class given data (Posterior Probability)
  - P(C) : Prior Probability
  - P(D|C) : Likelihood
  - P(D) : constant, so ignored

Does patient have cancer or not?

- A patient takes a lab test and the result comes back positive. It is known that the test returns a correct positive result in only 99% of the cases and a correct negative result in only 95% of t he cases. Furthermore, only 0.03 of the entire population has this disease.

1. What is the probability that this patient has cancer?
2. What is the probability that he does not have cancer?
3. What is the diagnosis?

Diamond card in the box?

- There is a deck of 52 randomly shuffled cards (13 spades, 13 diamonds, 13 hearts, and 13 clubs). We randomly pick one card out of the deck, without looking at its contents, and put it inside a box. Now, we randomly choose three cards out of the deck. We find that all of them are diamonds.
- What is the probability that the card in the box is diamond?

Two daughters?

- Suppose your neighbor has 2 children.

1. You ask to your neighbor if she has a daughter, and she replies “Yes”. What is the probability that the other child is also a girl?
2. You ask to your neighbor if her older child is a girl, and she replies “Yes”. What is the probability that the younger child is also a girl?
3. You meet one of the children, and it is a daughter. What is the probability that the other child is also a girl?

Prosecutor's fallacy

- A crime-scene DNA sample is compared against a database of 20,000 men. A match is found, that man is accused and at his trial, it is testified that the probability that two DNA profiles match by chance is only 1 in 10,000.
- Suppose none of the men in the database actually left the crime-scene DNA, then what is the chance of getting at least one match among the records?

Bayes Classifier

- Recall $P(C|X1,X2,...Xn) = P(C)P(X1,X2,...Xn|C)/P(X1,X2,...Xn)$
- → $P(C|X1,X2,...Xn)\ \alpha\ P(C)P(X1,X2,...Xn|C)$

- Maximum A Posteriori hypothesis
  - $hMAP = \text{argmax}\ P(D|h)P(h)$
- Maximum Likelihood hypothesis
  - $hML = \text{argmax}\ P(D|h)$

Naïve Bayes Classifier

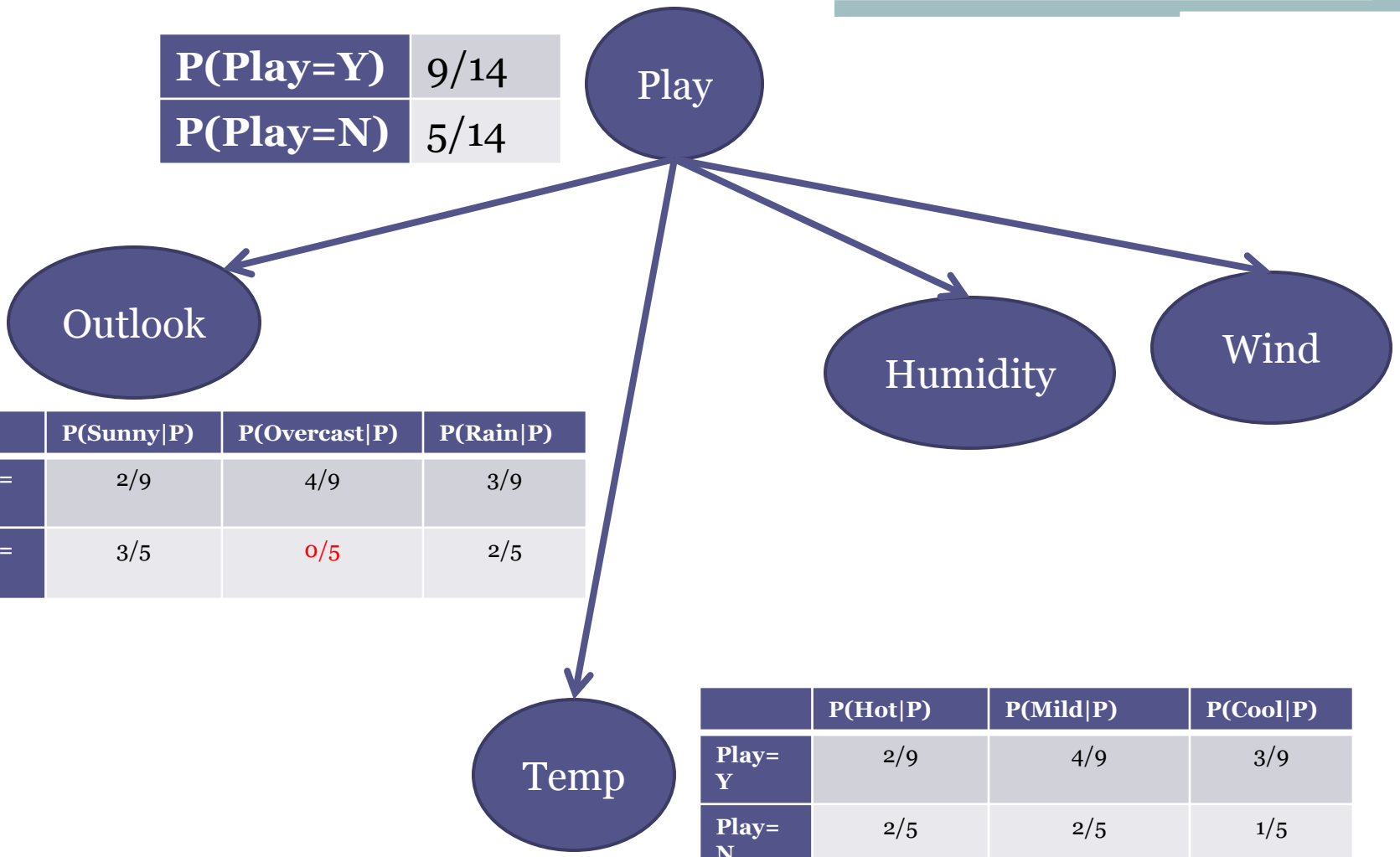
- Recall $P(C|X1,X2,...Xn) = P(C)P(X1,X2,...Xn|C)/P(X1,X2,...Xn)$
- → $P(C|X1,X2,...Xn) \alpha P(C)P(X1,X2,...Xn|C)$
- It is hard to calculate joint probabilities
  - Too many cases
  - Too little data
- If we ignore the dependences among X1,X2,...Xn?
  - Why? Because we are naïve.
  - More precisely, suppose X1,X2,...Xn are conditionally independent of each other given C
- $P(C|X1,X2,...Xn) \alpha P(C) * \Pi P(Xi|C)$

Naive Bayes Classifier Example

Day	Outlook	Temp	Humidty	Wind	Play?
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

Question: For the day <sunny, cool, high, strong>, what's the play prediction?

P(Play=Y)	9/14
P(Play=N)	5/14



	P(Sunny|P)	P(Overcast|P)	P(Rain|P)
Play=Y	2/9	4/9	3/9
Play=N	3/5	0/5	2/5

	P(Hot|P)	P(Mild|P)	P(Cool|P)
Play=Y	2/9	4/9	3/9
Play=N	2/5	2/5	1/5

Zero Occurrence

- When a feature is never occurred in the training set → zero frequency → PANIC: makes all terms zero
- Smoothing the distribution
  - Laplacian Smoothing
  - Dirichlet Priors Smoothing
  - and many more (Absolute Discouting, Jelinek-Mercer smoothing, Katz smoothing, Good-Turing smoothing, etc.)

Extension of NBC

- NBC is actually very effective

- Selective Bayesian Classifiers (SBC)
- Tree Augmented Naïve Bayes (TAN)
  - ChowLiu algorithm (CL-TAN)
  - SuperParent algorithm (SP-TAN)
- Attribute Value Taxonomy Guided Naïve Bayes Learner (AVT-NBL)
- n-gram Augmented Naïve Bayes