Introduction to Probability Theory 2

Clayton Greenberg
CoLi, CS, MMCI, LSV, CRC 1102 (IDeaL) B4

October 23, 2014

Schedule

22.10.2014	Calculate the probability of a given parse
23.10.2014	Solve the medical test Bayes' Rule problem
27.10.2014	Create a code for simplified Polynesian
29.10.2014	Identify types of machine learning problems
31.10.2014	Find a regression line for 2D data

From last time

1. probability = what you want / what is possible
2. “and” = * (times) [if independent]
3. “or” = + (plus) [if mutually exclusive]
4. logarithms = exponents
5. surprisal = the negative logarithm of probability

Conditional probability example

Definition of conditional probability

- p(**A** | **B**) = p(**A**, **B**) / p(**B**)
- conditional = joint / normalizer
- "How likely is **A**, given that **B** happens?"
- "Germany wins the game" is a *predicate* that describes event **A** (a set of outcomes).

Conditional probability and independence

- $p(\mathbf{A} \mid \mathbf{B}) = p(\mathbf{A}, \mathbf{B}) / p(\mathbf{B})$
- Independent: $p(\mathbf{A}, \mathbf{B}) = p(\mathbf{A}) * p(\mathbf{B})$
- Then, $p(\mathbf{A} \mid \mathbf{B}) = p(\mathbf{A}) * p(\mathbf{B}) / p(\mathbf{B})$
- $p(\mathbf{A} \mid \mathbf{B}) = p(\mathbf{A})$

Conditional independence

- $p(\mathbf{A}, \mathbf{B} \mid \mathbf{C}) = p(\mathbf{A} \mid \mathbf{C}) * p(\mathbf{B} \mid \mathbf{C})$
- Conditional independence does NOT imply independence.
- Independence does NOT imply conditional independence.

The chain rule

- p($\mathbf{A_3}$, $\mathbf{A_2}$, $\mathbf{A_1}$) = p($\mathbf{A_3}$ | $\mathbf{A_2}$, $\mathbf{A_1}$)*p($\mathbf{A_2}$ | $\mathbf{A_1}$)*p($\mathbf{A_1}$)
- p($\mathbf{A_n}$, **…**, $\mathbf{A_1}$) = p($\mathbf{A_n}$ | $\mathbf{A_{n-1}}$, …, $\mathbf{A_1}$)*p($\mathbf{A_{n-1}}$, …, $\mathbf{A_1}$)
- chain rule = the conditional of the last times the joint of the rest

$$\mathrm{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathrm{P}\left(A_k \,\middle|\, \bigcap_{j=1}^{k-1} A_j\right)$$

Probabilistic context free grammar (PCFG)

S → NP VP (1.0)

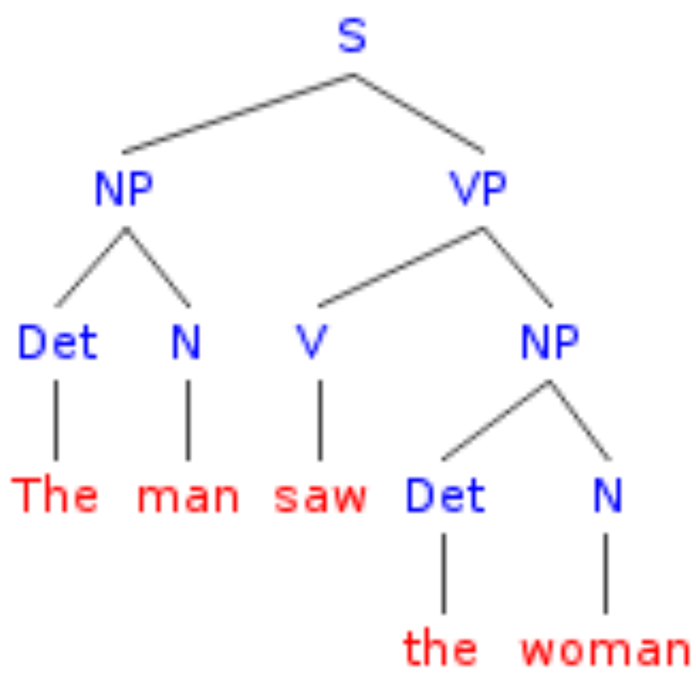
NP → Det N (0.8)

NP → NP PP (0.2)

PP → P NP (1.0)

VP → V NP (0.7)

VP → VP PP (0.3)



Probabilistic context free grammar (PCFG)

$p(NP, VP \mid S) = 1.0$

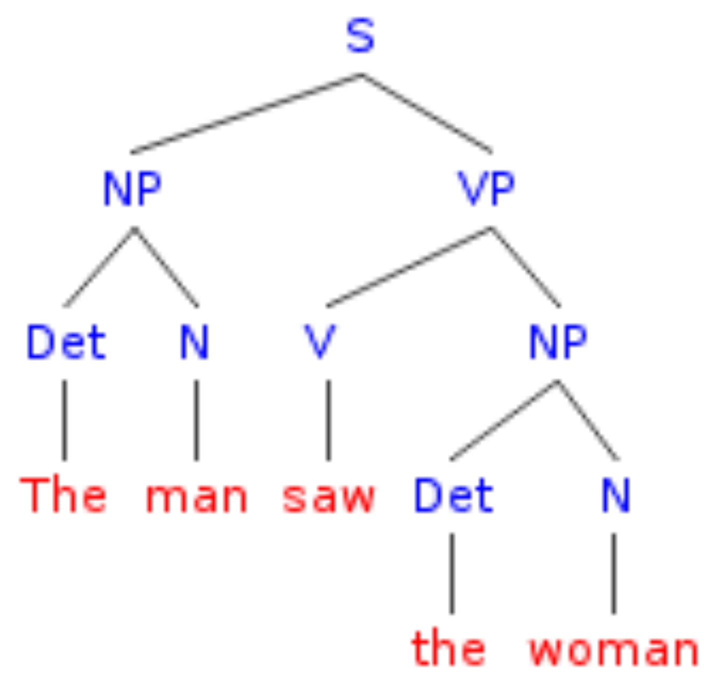
$p(Det, N \mid NP) = 0.8$

$p(NP, PP \mid NP) = 0.2$

$p(P, NP \mid PP) = 1.0$

$p(V, NP \mid VP) = 0.7$

$p(VP, PP \mid VP) = 0.3$



Probability of a string

p(“the man saw the woman” | S)

= p(“…”, NP, VP | S) [marginal probability sum]

= p(“…” | NP, VP, S)*p(NP, VP | S) [Chain Rule]

= p(“…” | NP, VP, S)*(1.0) [grammar rule]

= p(“…” | NP, VP)*(1.0) [independence assumption]

Then split [using independence assumption]:

= p(“the man” | NP)*p(“saw the woman” | VP)*(1.0)

…

Probability of a string

p("the man saw the woman" | S)

= p(the | Det)*p(man | N)*p(saw | V)*p(the | Det)*p(woman | N)*(1.0)(0.8)(0.7)(0.8)

These p(WORD | TAG) terms are *lexical rules* and are usually part of the PCFG.

Toy PCFG

S → NP VP (1.0)
NP → Det N (0.8)
NP → NP PP (0.2)
PP → P NP (1.0)
VP → V NP (0.7)
VP → VP PP (0.3)

Det → "the" (1.0)
N → "man" (0.5)
N → "woman" (0.5)
V → "saw" (1.0)
P → "with" (1.0)

Toy PCFG

$p(NP, VP \mid S) = 1.0$

$p(Det, N \mid NP) = 0.8$

$p(NP, PP \mid NP) = 0.2$

$p(P, NP \mid PP) = 1.0$

$p(V, NP \mid VP) = 0.7$

$p(VP, PP \mid VP) = 0.3$

$p(the \mid Det) = 1.0$

$p(man \mid N) = 0.5$

$p(woman \mid N) = 0.5$

$p(saw \mid V) = 1.0$

$p(with \mid P) = 1.0$

Probability of a string

p("the man saw the woman" | S)

= p(the | Det)*p(man | N)*p(saw | V)

*p(the | Det)*p(woman | N)*(1.0)(0.8)(0.7)(0.8)

probability of a string (PCFG) =

product of used rules (grammatical and lexical)

Probability of a string

p("the man saw the woman" | S)

= (1.0)(0.5)(1.0)(1.0)(0.5)(1.0)(0.8)(0.7)(0.8)

= 0.112

probability of a string (PCFG) =

product of used rules (grammatical and lexical)

Bayes' rule

- So, we have p("…" | S).
- What if we want p(S | "…")?
- By the Chain Rule, p(**A**, **B**) = p(**A** | **B**)*p(**B**)
- Also, by the Chain Rule, p(**A**, **B**) = p(**B** | **A**)*p(**A**)
- So, p(**A** | **B**)*p(**B**) = p(**A**, **B**) = p(**B** | **A**)*p(**A**)
- Or: p(**A** | **B**) = p(**B** | **A**)*p(**A**) / p(**B**)

Parts of Bayes' rule

p(**A** | **B**)	=	p(**B** | **A**)	*	p(**A**)	/	p(**B**)
posterior		likelihood	*	prior	/	normalizer

The most likely event

- Event alternatives: $\mathbf{A_1, A_2, A_3, \ldots, A_n}$
- Common context event: **B**
- $\text{argmax}_i \; p(\mathbf{A_i} \mid \mathbf{B}) = \text{argmax}_i \; p(\mathbf{B} \mid \mathbf{A_i}) * p(\mathbf{A_i}) / p(\mathbf{B})$

  $= \text{argmax}_i \; p(\mathbf{B} \mid \mathbf{A_i}) * p(\mathbf{A_i})$

Medical test Bayes' rule problem

- Rare disease: affects 1 in 10,000 people.
- If someone has the disease, s/he will test positive with probability 0.97.
- If someone does not have the disease, s/he will test positive with probability 0.01.
- You just tested positive for the disease.

1. What is your most likely health status?
2. What is the probability that you have the disease?

Medical test Bayes' rule problem

- p(disease) = 0.0001 [prior]
- p(positive | disease) = 0.97 [likelihood]
- p(positive | ¬disease) = 0.01 [for normalizer]

1. p(positive | disease) * p(disease) ? p(positive | ¬disease) * p(¬disease)
2. p(disease | positive) = p(positive | disease) * p(disease) / p(positive)

Exercises

1. Memorize the green statements:
   1. conditional = joint / normalizer
   2. chain rule: joint = conditional of last * joint of rest
   3. probability of a string (PCFG) = product of used rules
   4. Bayes' rule: p(**A** | **B**) = p(**B** | **A**) * p(**A**) / p(**B**)
   5. Bayes' rule: posterior = likelihood * prior / normalizer

2. Solve the medical test Bayes' Rule problem (slide 20).