For a more in-depth introduction to Naïve Bayes Classifiers and the theory surrounding them, please see Andrew's lecture on Probability for Data Miners.

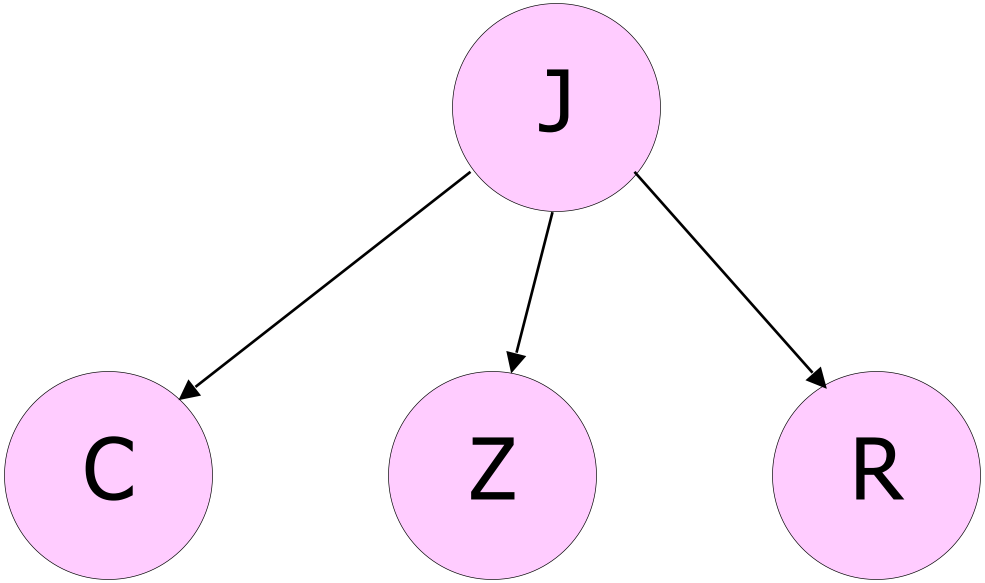
Naïve Bayes Classifiers

Note to other teachers and users of these slides. Andrew would be delighted if you found this source material useful in giving your own lectures. Feel free to use these slides verbatim, or to modify them to fit your own needs. PowerPoint originals are available. If you make use of a significant portion of these slides in your own lecture, please include this message, or the following link to the source repository of Andrew's tutorials: http://www.cs.cmu.edu/~awm/tutorials . Comments and corrections gratefully received.

Andrew W. Moore
Professor
School of Computer Science
Carnegie Mellon University
www.cs.cmu.edu/~awm
awm@cs.cmu.edu
412-268-7599

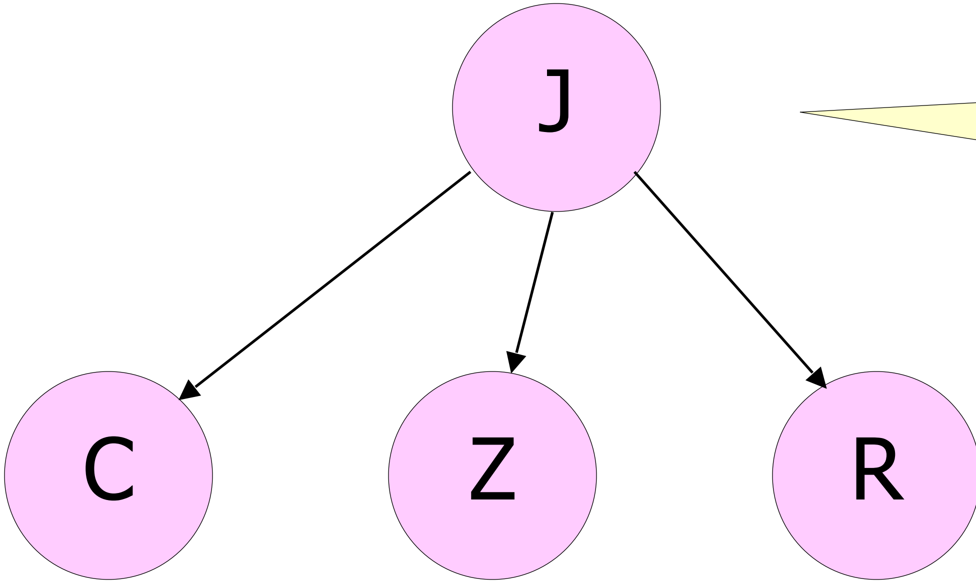
These notes assume you have already met Bayesian Networks

A simple Bayes Net



J	Person is a Junior
C	Brought Coat to Classroom
Z	Live in zipcode 15213
R	Saw "Return of the King" more than once

A simple Bayes Net

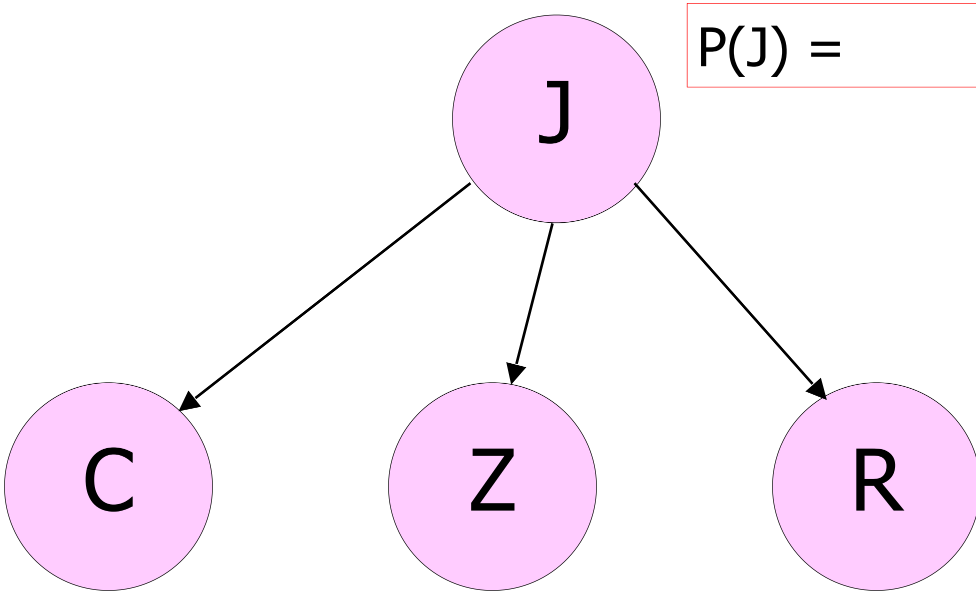


What parameters are stored in the CPTs of this Bayes Net?

J	Person is a Junior
C	Brought Coat to Classroom
Z	Live in zipcode 15213
R	Saw "Return of the King" more than once

A simple Bayes Net

P(J) =



J	Person is a Junior
C	Brought Coat to Classroom
Z	Live in zipcode 15213
R	Saw "Return of the King" more than once

$P(C|J) =$

$P(C|\sim J) =$

$P(Z|J) =$

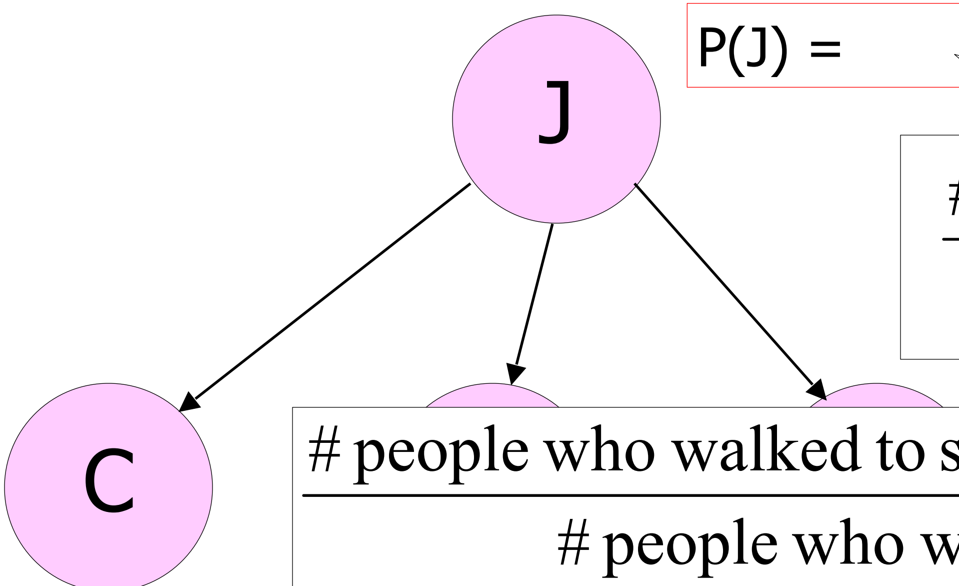
$P(Z|\sim J) =$

$P(R|J) =$

$P(R|\sim J) =$

Suppose we have a database from 20 people who attended a lecture. How could we use that to estimate the values in this CPT?

A simple Bayes Net



P(J) =

$$\frac{\text{\# people who walked to school}}{\text{\# people in database}}$$

Person is a Junior

than once

$$\frac{\text{\# people who walked to school and brought a coat}}{\text{\# people who walked to school}}$$

P(C|J) =

P(C|~J)=

P(Z|J) =

P(Z|~J)=

P(R|J) =

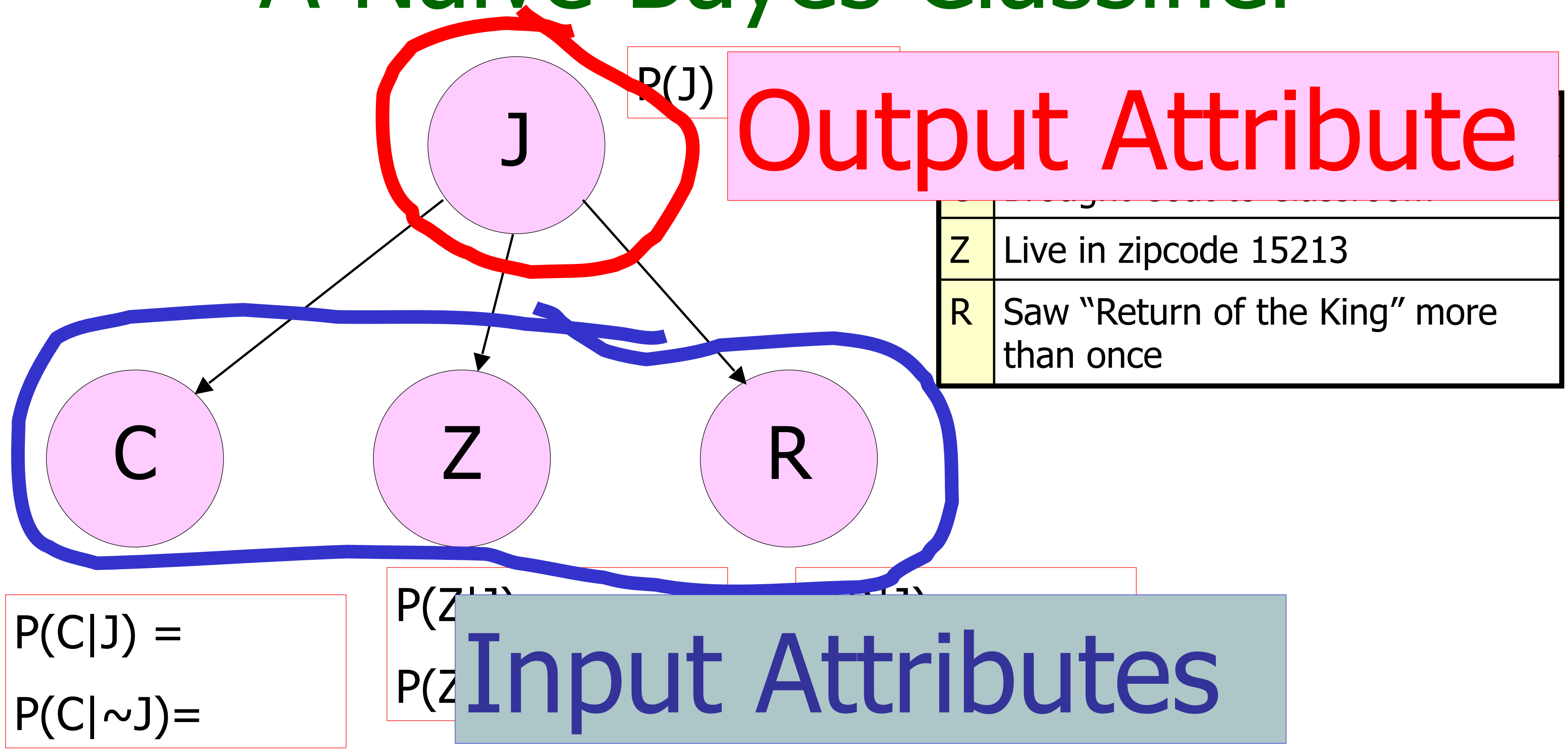
P(R|~J)=

$$\frac{\text{\# coat - bringers who didn't walk to school}}{\text{\# people who didn't walk to school}}$$

...abase from 20 people

...e that

A Naïve Bayes Classifier

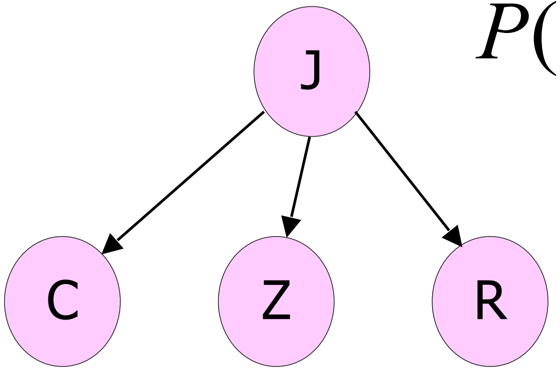


Z	Live in zipcode 15213
R	Saw "Return of the King" more than once

A new person shows up at class wearing an "I live right above the Manor Theater where I saw all the Lord of The Rings Movies every night" overcoat.

What is the probability that they are a Junior?

Naïve Bayes Classifier Inference



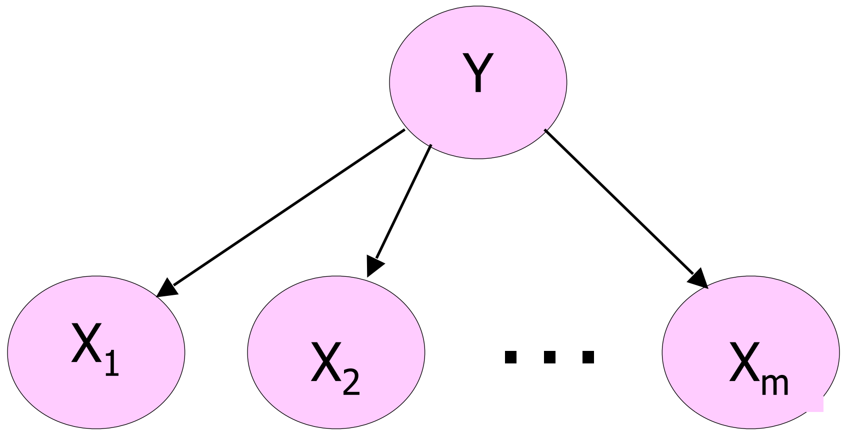
$$P(J \mid C \wedge \neg Z \wedge R) =$$

$$= \frac{P(J \wedge C \wedge \neg Z \wedge R)}{P(C \wedge \neg Z \wedge R)}$$

$$= \frac{P(J \wedge C \wedge \neg Z \wedge R)}{P(J \wedge C \wedge \neg Z \wedge R) + P(\neg J \wedge C \wedge \neg Z \wedge R)}$$

$$= \frac{P(C \mid J)P(\neg Z \mid J)P(R \mid J)P(J)}{\begin{pmatrix} P(C \mid J)P(\neg Z \mid J)P(R \mid J)P(J) \\ + \\ P(C \mid \neg J)P(\neg Z \mid \neg J)P(R \mid \neg J)P(\neg J) \end{pmatrix}}$$

The General Case



1. Estimate P(Y=v) as fraction of records with Y=v
2. Estimate P(X_i=u | Y=v) as fraction of "Y=v" records that also have X=u.
3. To predict the Y value given observations of all the X_i values, compute

$$Y^{\text{predict}} = \underset{v}{\operatorname{argmax}}\, P(Y = v \mid X_1 = u_1 \boxed{?} X_m = u_m)$$

Naïve Bayes Classifier

$$Y^{\text{predict}} = \operatorname*{argmax}_{v} P(Y = v \mid X_1 = u_1 \boxed{?} X_m = u_m)$$

$$Y^{\text{predict}} = \operatorname*{argmax}_{v} \frac{P(Y = v \wedge X_1 = u_1 \boxed{?} X_m = u_m)}{P(X_1 = u_1 \boxed{?} X_m = u_m)}$$

$$Y^{\text{predict}} = \operatorname*{argmax}_{v} \frac{P(X_1 = u_1 \boxed{?} X_m = u_m \mid Y = v) P(Y = v)}{P(X_1 = u_1 \boxed{?} X_m = u_m)}$$

$$Y^{\text{predict}} = \operatorname*{argmax}_{v} P(X_1 = u_1 \boxed{?} X_m = u_m \mid Y = v) P(Y = v)$$

Because of the structure of the Bayes Net

$$Y^{\text{predict}} = \operatorname*{argmax}_{v} P(Y = v) \prod_{j=1}^{n_Y} P(X_j = u_j \mid Y = v)$$

More Facts About Naïve Bayes Classifiers

- Naïve Bayes Classifiers can be built with real-valued inputs*
- Rather Technical Complaint: Bayes Classifiers don't try to be maximally discriminative---they merely try to honestly model what's going on*
- Zero probabilities are painful for Joint and Naïve. A hack (justifiable with the magic words "Dirichlet Prior") can help*.
- Naïve Bayes is wonderfully cheap. And survives 10,000 attributes cheerfully!

*See future Andrew Lectures

What you should know

- How to build a Bayes Classifier
- How to predict with a BC