Bayesian Networks
(aka Bayes Nets, Belief Nets)
(one type of Graphical Model)

[based on slides by Jerry Zhu and Andrew Moore]

Full Joint Probability Distribution
Making a joint distribution of $N$ variables:
1. List all combinations of values (if each variable has $k$ values, there are $k^N$ combinations)
2. Assign each combination a probability
3. They should sum to 1

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>150/365</td>
</tr>
<tr>
<td>Sunny</td>
<td>Cold</td>
<td>50/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Hot</td>
<td>40/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Hot</td>
<td>5/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
</tbody>
</table>

Using the Full Joint Distribution
- Once you have the joint distribution, you can do anything, e.g. marginalization:
  \[ P(E) = \sum_{\text{rows matching } E} P(\text{row}) \]
- e.g., \( P(\text{Sunny or Hot}) = (150+50+40+5)/365 \)
  Convince yourself this is the same as \( P(\text{sunny}) + P(\text{hot}) - P(\text{sunny and hot}) \)

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>150/365</td>
</tr>
<tr>
<td>Sunny</td>
<td>Cold</td>
<td>50/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Hot</td>
<td>40/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Hot</td>
<td>5/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
</tbody>
</table>

Using the Joint Distribution
- You can also do inference:
  \[ P(Q \mid E) = \frac{\sum_{\text{rows matching } E \text{ AND } E} P(\text{row})}{\sum_{\text{rows matching } E} P(\text{row})} \]
  \( P(\text{Hot} \mid \text{Rainy}) \)

<table>
<thead>
<tr>
<th>Weather</th>
<th>Temperature</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>Hot</td>
<td>150/365</td>
</tr>
<tr>
<td>Sunny</td>
<td>Cold</td>
<td>50/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Hot</td>
<td>40/365</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Hot</td>
<td>5/365</td>
</tr>
<tr>
<td>Rainy</td>
<td>Cold</td>
<td>60/365</td>
</tr>
</tbody>
</table>
**The Bad News**

- Joint distribution requires a lot of storage space
- For $N$ variables, each taking $k$ values, the joint distribution has $k^N$ numbers (and $k^N - 1$ degrees of freedom)
- It would be nice to use fewer numbers …
- Bayesian Networks to the rescue!
  - Provides a decomposed representation of the FJPD
  - Encodes a collection of conditional independence relations

---

**Introducing Bayesian Networks**

| $P(B)$ | $P(A | B, E)$ | $P(A | B, \sim E)$ |
|--------|---------------|-------------------|
| 0.001  | 0.95          | 0.94              |
| 0.002  | 0.95          | 0.94              |

| $P(B | E)$ | $P(A | B)$ |
|----------|-----------|
| 0.001    | 0.001     |

DAG, arcs often direct causation, but don’t have to be!

---

**Joint Probability from Bayes Net**

$$P(x_1, \ldots x_N) = \prod_i P(x_i | \text{parents}(x_i))$$

- Example: $P(\sim B, E, \sim A) = P(\sim B) P(E) P(\sim A | \sim B, E)$

- DAG, arcs often direct causation, but don’t have to be!

---

**With Bayes Net**

| $P(A | B, E)$ | $P(A | B, \sim E)$ | $P(A | \sim B, E)$ | $P(A | \sim B, \sim E)$ |
|--------------|-------------------|-------------------|-----------------------|
| 0.95         | 0.94              | 0.29              | 0.001                 |

Recall the chain rule:

$$P(\sim B, E, \sim A) = P(\sim B) P(E | \sim B) P(\sim A | \sim B, E)$$

DAG, arcs often direct causation, but don’t have to be!
Bayesian Networks

- Directed, acyclic graphs (DAGs)
- Nodes = random variables
  - CPT stored at each node quantifies conditional probability of node’s r.v. given all its parents
- Arc from A to B means A "has a direct influence on" or "causes" B
  - Evidence for A increases likelihood of B (deductive influence from causes to effects)
  - Evidence for B increases likelihood of A (abductive influence from effects to causes)
- Encodes conditional independence assumptions

Example

- A: your alarm sounds
- J: your neighbor John calls you
- M: your other neighbor Mary calls you
- John and Mary do not communicate (they promised to call you whenever they hear the alarm)

What kind of independence do we have?
- **Conditional independence:** \( P(J, M | A) = P(J | A) P(M | A) \)

What does the Bayes Net look like?

Our BN: \( P(A, J, M) = P(A) P(J | A) P(M | A) \)

Chain rule: \( P(A, J, M) = P(A) P(J | A) P(M | A, J) \)

Our B.N. assumes conditional independence, so \( P(M | A, J) = P(M | A) \)

What kind of independence do we have?
- **Conditional independence** \( P(J, M | A) = P(J | A) P(M | A) \)

What does the Bayes Net look like?
**A Simple Bayesian Network**

\[
S \in \{ \text{no, light, heavy} \} \\
C \in \{ \text{none, benign, malignant} \}
\]

\[
P(S=\text{no}) = 0.80 \\
P(S=\text{light}) = 0.15 \\
P(S=\text{heavy}) = 0.05
\]

<table>
<thead>
<tr>
<th></th>
<th>no</th>
<th>light</th>
<th>heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(C=\text{none}</td>
<td>S=))</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>(P(C=\text{benign}</td>
<td>S=))</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>(P(C=\text{malignant}</td>
<td>S=))</td>
<td>0.01</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**A Bayesian Network**

**Diagnostic variables**

Flu

Allergy

Sinus

Runny Nose

Headache

**Evidence variables**

**A Simple Bayesian Network**

\[
S \in \{ \text{no, light, heavy} \} \\
C \in \{ \text{none, benign, malignant} \}
\]

\[
P(S=\text{no}) = 0.80 \\
P(S=\text{light}) = 0.15 \\
P(S=\text{heavy}) = 0.05
\]

<table>
<thead>
<tr>
<th></th>
<th>no</th>
<th>light</th>
<th>heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P(C=\text{none}</td>
<td>S=))</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>(P(C=\text{benign}</td>
<td>S=))</td>
<td>0.03</td>
<td>0.08</td>
</tr>
<tr>
<td>(P(C=\text{malignant}</td>
<td>S=))</td>
<td>0.01</td>
<td>0.04</td>
</tr>
</tbody>
</table>

**A Bayesian Network**

Age

Gender

Exposure to Toxics

Smoking

Cancer

Serum Calcium

Lung Tumor

Not needed
Applications

- Medical diagnosis systems
- Manufacturing system diagnosis
- Computer systems diagnosis
- Network systems diagnosis
- Helpdesk troubleshooting
- Information retrieval
- Customer modeling

RICOH Fixit

- Diagnostics and information retrieval

FIXIT: Ricoh copy machine

Online Troubleshooters
**Pathfinder**

- Pathfinder is one of the first BN systems
- It performs diagnosis of lymph-node diseases
- It deals with over 60 diseases and 100 symptoms and test results
- 14,000 probabilities
- Commercialized by Intellipath and Chapman Hall and applied to about 20 tissue types

**Example: Car insurance**

**Example: Car diagnosis**

Initial evidence: car won’t start
Testable variables (green), “broken, so fix it” variables (orange)
Hidden variables (gray) ensure sparse structure, reduce parameters
Conditional Independence in Bayes Nets

- A node is conditionally independent of its non-descendents, given its parents.
- A node is conditionally independent of all other nodes, given its "Markov blanket" (i.e., parents, children, and children's parents).

Conditional Independence

\[ \text{Cancer} \text{ is conditionally independent of Age and Gender given Smoking} \]

More Conditional Independence

- \( P(L \mid SC, C) = P(L \mid C) \)

Interpreting Bayesian Nets

- 2 nodes are unconditionally independent if there’s no undirected path between them.
- If there’s an undirected path between 2 nodes, then whether or not they are independent or dependent depends on what other evidence is known.

A and B are independent given nothing else, but are dependent given C.
Example with 5 Variables

- B: there's burglary in your house
- E: there's an earthquake
- A: your alarm sounds
- J: your neighbor John calls you
- M: your other neighbor Mary calls you

- B, E are independent
- J is directly influenced by only A (i.e., J is conditionally independent of B, E, M, given A)
- M is directly influenced by only A (i.e., M is conditionally independent of B, E, J, given A)

Creating a Bayes Net

- **Step 1:** add variables. Choose the variables you want to include in the Bayes Net

```
B  E
A  J
M
```

- **Step 2:** add directed edges
  - The graph must be acyclic
  - If node X is given parents Q₁, ..., Qᵢ, you are promising that any variable that's not a descendant of X is conditionally independent of X given Q₁, ..., Qᵢ

```
B  E
A
J  M
```

- **Step 3:** add CPT's
  - Each table must list \( P(X \mid \text{Parent values}) \) for all combinations of parent values
  - e.g. you must specify \( P(J \mid A) \) AND \( P(J \mid \neg A) \). They don't have to sum to 1!

```
P(B) = 0.001
P(A \mid B, E) = 0.95
P(A \mid B, \neg E) = 0.94
P(A \mid \neg B, E) = 0.29
P(A \mid \neg B, \neg E) = 0.001
```

```
P(E) = 0.002
P(A) = 0.7
P(M \mid A) = 0.01
P(M \mid \neg A) = 0.05
```

```
P(J \mid A) = 0.9
P(J \mid \neg A) = 0.05
```

```xml
B: there's burglary in your house
E: there's an earthquake
A: your alarm sounds
J: your neighbor John calls you
M: your other neighbor Mary calls you
```
Creating a Bayes Net

1. Choose a set of relevant variables
2. Choose an ordering of them, call them $x_1, \ldots, x_N$
3. for $i = 1$ to $N$:
   1. Add node $x_i$ to the graph
   2. Set parents($x_i$) to be the minimal subset of
      $\{x_1, \ldots, x_{i-1}\}$, such that $x_i$ is conditionally
      independent of all other members of $\{x_1, \ldots, x_{i-1}\}$
      given parents($x_i$)
3. Define the CPT's for $P(x_i \mid \text{assignments of parents}(x_i))$

   • Different ordering leads to different graph, in general
   • Best ordering when each var is considered after all vars that
     directly influence it

Compactness of Bayes Nets

• A Bayesian Network is a graph structure for
  representing conditional independence relations in a
  compact way
• A Bayes net encodes a joint distribution, often with
  far less parameters (i.e., numbers)
• A full joint table needs $k^N$ parameters ($N$ variables, $k$ values per variable)
  • grows exponentially with $N$
• If the Bayes net is sparse, e.g., each node has at
  most $M$ parents ($M \ll N$), only needs $O(Nk^M)$
  • grows linearly with $N$
  • can’t have too many parents, though

Computing a Joint Entry from a Bayes Net

How to compute an entry in the joint distribution?
E.g., what is $P(S, \neg M, L, \neg R, T)$?

Computing with Bayes Net

Apply the Chain Rule!
The General Case

\[
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) =
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) \times P(X_1=x_1, X_2=x_2, X_2=x_2, X_2=x_2)\]

\[
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) =
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) \times P(X_1=x_1, X_2=x_2, X_2=x_2)\]

\[
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) =
P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n) \times P(X_1=x_1, X_2=x_2)\]

Where are we Now?

- We defined a Bayes net, using small number of
  parameters, to describe the joint

- Any joint probability can be computed as

\[ P(x_1, \ldots, x_n) = \prod_i P(x_i | \text{parents}(x_i)) \]

- The above joint probability can be computed in time

- With this joint distribution, we can compute any
  conditional probability, \( P(Q | E) \), thus we can perform
  any inference

- How?

Computing Joint Probabilities using a Bayesian Network

How is any joint probability computed?

Sum the relevant joint probabilities:

Compute: \( P(a,b) \)

\[ = P(a,b,c,d) + P(a,b,c,\neg d) + P(a,b,\neg c,d) + P(a,b,\neg c,\neg d) \]

Compute: \( P(c) \)

\[ = P(a,b,c,d) + P(a,\neg b,c,d) + P(a,b,c,d) + P(\neg a,\neg b,c,d) + \]

- A BN can answer any query (i.e., probability) about the
  domain by summing the relevant joint probabilities

Inference by Enumeration

\[
P(Q | E) = \frac{\sum_{\text{joint matching } Q \text{ AND } E} P(\text{joint})}{\sum_{\text{joint matching } E} P(\text{joint})}
\]

by def. of cond. prob.

For example: \( P(B | J, \neg M) \)

1. Compute \( P(B,J,\neg M) \)
2. Compute \( P(J, \neg M) \)
3. Return \( P(B,J,\neg M)/P(J,\neg M) \)
Inference by Enumeration

\[
P(Q | E) = \frac{\sum_{\text{joint matching } Q \text{ AND } E} P(\text{joint})}{\sum_{\text{joint matching } E} P(\text{joint})}
\]

For example: \( P(B | J, ~M) \)

1. Compute \( P(B, J, ~M) \)
2. Compute \( P(J, ~M) \)
3. Return \( P(B, J, ~M) / P(J, ~M) \)

In general, if there are \( N \) variables, while evidence contains \( j \) variables, how many joints to sum up?

Another Example

Compute \( P(R \mid T, ~S) \) from the following Bayes Net

[Diagram showing the Bayes net with variables S, M, L, T, and R with their conditional probabilities.]
Another Example

Step 1: Compute $P(R, T, ~S)$
Step 2: Compute $P(T, ~S)$
Step 3: Return

$P(R, T, ~S)$

--------
$P(T, ~S)$

$P(M) = 0.6$

Compute $P(R \mid T, ~S)$?

Inference through a Bayes Net can go both “forward” and “backward” through arcs

• Causal (top-down) inference
  - Given a cause, infer its effects
  - E.g., $P(T \mid S)$

• Diagnostic (bottom-up) inference
  - Given effects/symptoms, infer a cause
  - E.g., $P(S \mid T)$
The Good News

We can do inference. That is, we can compute any conditional probability:

\[ P(\text{Some variable} \mid \text{Some other variable values}) \]

\[ P(E_1 \mid E_2) = \frac{P(E_1 \land E_2)}{P(E_2)} = \frac{\sum \text{joint entries matching } E_1 \text{ and } E_2}{\sum \text{joint entries matching } E_2} \]

“\text{Inference by Enumeration}” Algorithm

Parameter (CPT) Learning for BN

- Where do you get these CPT numbers?
  - Ask domain experts, or
  - Learn from data

\begin{align*}
P(B) &= 0.001 \\
P(A \mid B, E) &= 0.95 \\
P(A \mid B, \neg E) &= 0.94 \\
P(A \mid \neg B, E) &= 0.29 \\
P(A \mid \neg B, \neg E) &= 0.001 \\
P(E) &= 0.002 \\
P(J \mid A) &= 0.9 \\
P(J \mid \neg A) &= 0.05 \\
P(M \mid A) &= 0.7 \\
P(M \mid \neg A) &= 0.01
\end{align*}

Parameter (CPT) Learning for BN

- Learn from a data set like this:

\begin{align*}
&(-B, \neg E, -A, J, -M) \\
&(-B, -E, A, J, -M) \\
&(-B, -E, -A, J, -M) \\
&(-B, -E, -A, \neg J, -M) \\
&(-B, -E, A, \neg J, -M) \\
&(-B, E, A, J, M) \\
&(-B, E, -A, -J, M) \\
&(-B, E, -A, \neg J, M) \\
&(-B, E, A, \neg J, M) \\
&(-B, \neg E, A, J, M) \\
&(-B, \neg E, \neg A, J, M) \\
&(-B, \neg E, \neg A, \neg J, M) \\
&(-B, \neg E, A, J, M)
\end{align*}

How to learn this CPT?

\begin{align*}
P(B) &= 0.001 \\
P(A \mid B, E) &= 0.95 \\
P(A \mid B, \neg E) &= 0.94 \\
P(A \mid \neg B, E) &= 0.29 \\
P(A \mid \neg B, \neg E) &= 0.001 \\
P(E) &= 0.002 \\
P(J \mid A) &= 0.9 \\
P(J \mid \neg A) &= 0.05 \\
P(M \mid A) &= 0.7 \\
P(M \mid \neg A) &= 0.01
\end{align*}
Parameter (CPT) Learning for BN

- Learn from a data set like this:

\[ P(B) = \frac{\#(B)}{\#(B) + \#(\neg B)} \]

\[ P(A | B, E) = \frac{\#(A) \cap (B \land E)}{\#(B \land E)} \]

\[ P(A | B, \neg E) = \frac{\#(A) \cap (B \land \neg E)}{\#(B \land \neg E)} \]

\[ P(A | \neg B, E) = \frac{\#(A) \cap (\neg B \land E)}{\#(\neg B \land E)} \]

\[ P(A | \neg B, \neg E) = \frac{\#(A) \cap (\neg B \land \neg E)}{\#(\neg B \land \neg E)} \]

\[ P(E) = \frac{\#(E)}{\#(E) + \#(\neg E)} \]

\[ P(J | A) = \frac{\#(J \cap A)}{\#(A)} \]

\[ P(J | \neg A) = \frac{\#(J \cap \neg A)}{\#(\neg A)} \]

\[ P(M | A) = \frac{\#(M \cap A)}{\#(A)} \]

\[ P(M | \neg A) = \frac{\#(M \cap \neg A)}{\#(\neg A)} \]
Parameter (CPT) Learning for BN

- Learn from a data set like this:

\[
\begin{align*}
B & = 0.001 \\
P(A | B, E) & = 0.95 \\
P(A | B, \neg E) & = 0.94 \\
P(A | \neg B, E) & = 0.29 \\
P(A | \neg B, \neg E) & = 0.001 \\
P(E) & = 0.002 \\
P(J | A) & = 0.9 \\
P(J | \neg A) & = 0.05 \\
P(M | A) & = 0.7 \\
P(M | \neg A) & = 0.01
\end{align*}
\]

... Count \#(A) and \#(\neg A) in dataset where \(B = \text{false}\) and \(E = \text{true}\).

\[
P(A | B, E) = \frac{\#(A)}{\#(A) + \#(\neg A)}
\]

... Count \#(A) and \#(\neg A) in dataset where \(B = \text{false}\) and \(E = \text{false}\).

\[
P(A | B, \neg E) = \frac{\#(A)}{\#(A) + \#(\neg A)}
\]

'Unseen event' problem

- 'Unseen event' problem

\[
\begin{align*}
B & = 0.001 \\
P(A | B, E) & = 0.95 \\
P(A | B, \neg E) & = 0.94 \\
P(A | \neg B, E) & = 0.29 \\
P(A | \neg B, \neg E) & = 0.001 \\
P(E) & = 0.002 \\
P(J | A) & = 0.9 \\
P(J | \neg A) & = 0.05 \\
P(M | A) & = 0.7 \\
P(M | \neg A) & = 0.01
\end{align*}
\]

... Count \#(A) and \#(\neg A) in dataset where \(B = \text{true}\) and \(E = \text{true}\).

\[
P(A | B, E) = \frac{\#(A)}{\#(A) + \#(\neg A)}
\]

... Count \#(A) and \#(\neg A) in dataset where \(B = \text{true}\) and \(E = \text{false}\).

\[
P(A | B, \neg E) = \frac{\#(A)}{\#(A) + \#(\neg A)}
\]

What if there's no row with \((B, E, \neg A, *, *)\) in the dataset?

Do you want to set

\[
P(A | B, E) = 1 \\
P(A | \neg B, \neg E) = 0?
\]

Why or why not?

Parameter (CPT) Learning for BN

\[
P(X=x | \text{parents}(X)) = (\text{frequency of } x \text{ given parents})
\]

is called the Maximum Likelihood (ML) estimate

ML estimate is vulnerable to 'unseen event' problem when dataset is small

- flip a coin 3 times, all heads \(\rightarrow\) one-sided coin?

Simplest solution: 'Add one' smoothing
Smoothing CPT

- ‘Add one’ smoothing: **add 1 to all counts**
- In the previous example, count #(#A) and #(#~A) in dataset where B=true and E=true
  - \[ P(A|B,E) = \frac{[\#(A)+1]}{[\#(A)+1] + \#(~A)+1} \]
  - If #(A)=1, #(~A)=0:
    - without smoothing \( P(A|B,E) = 1, P(~A|B,E) = 0 \)
    - with smoothing \( P(A|B,E) = 0.67, P(~A|B,E) = 0.33 \)
  - If #(A)=100, #(~A)=0:
    - without smoothing \( P(A|B,E) = 1, P(~A|B,E) = 0 \)
    - with smoothing \( P(A|B,E) = 0.99, P(~A|B,E) = 0.01 \)
- Smoothing bravely saves you when you don’t have enough data, and humbly hides away when you do
- It’s a form of Maximum a posteriori (MAP) estimation

---

Naïve Bayes Classifier

- Find \( v = \arg\max_v P(Y = v) \prod_{i=1}^n P(X_i = u_i | Y = v) \)
  - Class variable
  - Evidence variable

- Assumes all evidence variables are conditionally independent of each other given the class variable
- Robust since it gives the right answer as long as the correct class is more likely than all others

---

BN Special Case: Naïve Bayes

- A special Bayes Net structure:
  - a ‘class’ variable \( Y \) at root, compute \( P(Y | X_1, \ldots, X_N) \)
  - evidence nodes \( X_i \) (observed features) are all leaves
  - conditional independence between all evidence assumed. Usually not valid, but often empirically OK

---

A Special BN: Naïve Bayes Classifiers

- What’s stored in the CPTs?
A Special BN: Naïve Bayes Classifiers

\[ P(J) = \]

\[
\begin{array}{c|c}
\text{J} & \text{Person is Junior} \\
\text{C} & \text{Brought coat to class} \\
\text{Z} & \text{Lives in zipcode 53706} \\
\text{H} & \text{Saw "Hunger Games" more than once}
\end{array}
\]

\[ P(C|J) = \]

\[ P(Z|J) = \]

\[ P(H|J) = \]

\[ P(C|\sim J) = \]

\[ P(Z|\sim J) = \]

\[ P(H|\sim J) = \]

Is the Person a Junior?

- Input (evidence): C, Z, H
- Output (query): J

\[
P(J|C,Z,H) = \frac{P(J,C,Z,H)}{P(C,Z,H)} \quad \text{by def. of cond. prob.}
\]

\]

where

\[
P(J,C,Z,H) = P(J)P(C|J)P(Z|J)P(H|J) \quad \text{by chain rule and}
\]

conditional independence associated with B.N.

\[
\]

Application: Bayesian Networks for Breast Cancer Diagnosis

Breast Cancer

Abnormal mammo

Elizabeth S. Burnside
Department of Radiology
University of Wisconsin Hospitals
What You Should Know

- Inference with joint distribution
- Problems of joint distribution
- Bayes Net: representation (nodes, edges, CPT) and meaning
- Compute joint probabilities from Bayes net
- Inference by enumeration
- Naïve Bayes

Results

<table>
<thead>
<tr>
<th></th>
<th>Radiologist</th>
<th>Bayes Net</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.916</td>
<td>.919</td>
<td>.948</td>
</tr>
</tbody>
</table>

ROC curves

- FPF: False Positive Fraction
- TPF: True Positive Fraction

What You Should Know