From Data To Understanding

• In machine learning, maintain critical perspective
  ▪ Making predictions is only part of the story
  ▪ Also try to get some **understanding** of the domain

• Example
  ▪ True statement: palm size negatively correlates with life expectancy
    ○ The larger your palm size, the shorter your life (on average)
  ▪ Why?
    ○ Women have smaller palms than men on average
    ○ Women live 5 years longer than men on average
  ▪ Sometimes you need better model of your domain!
Bayesian Networks

- Bayesian Network
  - Graphical representation of joint probability distribution
    - Node: random variable
    - Arc (X, Y): variable X has direct influence on variable Y
      - Call X a “parent” of Y
    - Each node X has conditional probability: P(X | parents(X))
    - Graph has no cycles (loops by following arcs)
      - Called “Directed Acyclic Graph” (DAG)
Network Shows Conditional Independence

- Conditional independence encoded in network
  - Each node (variable) is conditionally independent of its non-descendants, given its parents
  - In network above, Palm Size (PS) and Life Expectancy (LE) are conditionally independent, given Gender (G)
    - Formally: \( P(PS, LE | G) = P(PS | G) P(LE | G) \)
- Network structure provides insight about domain
Each node has conditional probability table (CPT)

- For node X: \( P(X | \text{Parents}(X)) \)
- Conditional independence modularizes joint probability:

\[
P(X_1, X_2, ..., X_m) = \prod_{i=1}^{m} P(X_i | \text{Parents}(X_i))
\]
Each node has conditional probability table (CPT)

- Reduces number of parameters needed in model
- Normally, need $2 \times 3 \times 3 - 1 = 18 - 1 = 17$ parameters
- Here, need $(2 - 1) + (6 - 2) + (6 - 2) = 9$ parameters
Bayesian Network for Naïve Bayes

- Welcome back, Naïve Bayes…
  - Now with new and improved “Bayesian Network” flavor!

  - Network structure encodes assumption:
    \[
    P(X \mid Y) = P(X_1, X_2, \ldots, X_m \mid Y) = \prod_{i=1}^{m} P(X_i \mid Y)
    \]

  - Full joint distribution can be computed as:
    \[
    P(X, Y) = P(Y)P(X \mid Y) = P(Y)\prod_{i=1}^{m} P(X_i \mid Y)
    \]
“Evidence” in Bayesian Networks

• In many machine learning examples:
  ▪ We observe all $X_1, X_2, \ldots, X_m$ input variables and predict single output variable $Y$

• In general case of probabilistic inference:
  ▪ Have a set of random variables $X_1, X_2, \ldots, X_m$
  ▪ **Subset** of the variables $X_1, X_2, \ldots, X_m$ are observed
    ○ Call observed variables $E_1, E_2, \ldots, E_k$ (E for “evidence”)
  ▪ Want to determine probability of some set of *unobserved* variables given the observed evidence
    ○ Call unobserved variables we care about $Y_1, Y_2, \ldots, Y_c$
  ▪ Formally, want: $P(Y_1, Y_2, \ldots, Y_c \mid E_1, E_2, \ldots, E_k)$
Evaluation of Evidence

- Consider the following Bayes Net:

\[
\begin{align*}
P(B = T | E = T) &= 0.1 \\
P(B = T | E = F) &= 0.6
\end{align*}
\]

\[
\begin{array}{c}
\text{Bloodshot Eyes (B)} \\
\text{Enough Sleep (E)} \\
\text{Got A on CS109 final (A)}
\end{array}
\]

\[
\begin{align*}
P(E = T) &= 0.7 \\
P(A = T | E = T, R = T) &= 0.8 \\
P(A = T | E = F, R = T) &= 0.7 \\
P(A = T | E = T, R = F) &= 0.4 \\
P(A = T | E = F, R = F) &= 0.2 \\
P(R = T) &= 0.5
\end{align*}
\]

- Determine \( P(A = T | B = T, R = T) \)
- Sum over unseen variables:

\[
P(A = T | B = T, R = T) = \frac{P(A = T, B = T, R = T)}{P(B = T, R = T)} = \frac{\sum_{E=T,F} P(A = T, B = T, R = T, E)}{\sum_{E=T,F} \sum_{A=T,F} P(B = T, R = T, E, A)}
\]
Evaluation of Evidence

- Consider the following Bayes Net:

\[ P(B = T \mid E = T) = 0.1 \]
\[ P(B = T \mid E = F) = 0.6 \]
\[ P(E = T) = 0.7 \]
\[ P(A = T \mid E = T, R = T) = 0.8 \]
\[ P(A = T \mid E = F, R = T) = 0.7 \]
\[ P(A = T \mid E = T, R = F) = 0.4 \]
\[ P(A = T \mid E = F, R = F) = 0.2 \]
\[ P(R = T) = 0.5 \]

- Determine \( P(A = T \mid B = T, R = T) \)

- Note that joint probability decomposes as:
  \[ P(A, B, E, R) = P(E)P(B \mid E)P(R)P(A \mid E, R) \]

- Plug in values from CPTs to compute joint probabilities
Probability Tree

- Model outcomes of probabilistic events with tree

- Useful for modeling decisions

  - Expected payoff: yes = $p(1,000,000 - 1) + (1 - p)(-1)$
  - no = 0
Let’s Play a Game

• Which choice would you make?

Play?

- yes
  - 0.5
  - $20
- no
  - $X ← “Certain Equivalent” (CE)

• For what value of X are you indifferent to playing?
  - X = 3
  - X = 7
  - X = 9
  - X = 10

• Certain equivalent is value of game to you
Utility

- Utility $U(x)$ is “value” you derive from $x$

- Can be monetary, but often includes intangibles
  - E.g., quality of life, life expectancy, personal beliefs, etc.

---

**Decision Tree**

- Play? [Yes/No]
  - Yes [0.5]
    - $20,000
  - No [0.5]
    - $10,000

- Play? [Yes/No]
  - Yes [0.5]
    - $U(20,000)$
  - No [0.5]
    - $U(0)$

- Play? [Yes/No]
  - Yes [0.5]
    - $U(10,000)$
Utility Curves

- Utility curve determines your “risk preference”
  - Can be different in different parts of the curve
Non-Linear Utility of Money

- These two choices are different for most people

\[
\begin{align*}
\text{Play?} & \quad 0.5 \quad \text{yes} \quad 0.5 \quad \text{no} \\
\text{yes} & \quad 0.5 \quad \$10 \\
\text{no} & \quad 0.5 \quad \$2 \\
\text{Play?} & \quad 0.5 \quad \text{yes} \quad 0.5 \quad \text{no} \\
\text{yes} & \quad 0.5 \quad \$100,000,000 \\
\text{no} & \quad 0.5 \quad \$20,000,000
\end{align*}
\]
Risk Premium

- A slightly different game:

- Expected monetary value (EMV) = expected dollar value of game (here = $10,000)

- Risk premium = EMV – CE = $3,000
  - How much you would pay (give up) to avoid risk
  - This is what insurance is all about

![Decision tree diagram]

- Say this is our CE

- Insure car?
  - no 0.02 $-30,000
  - yes 0.98 $-1000

- $-600
Exponential Utility Curves

- Many people have exponential utility curves

\[ U(x) = 1 - e^{-x/R} \]

- R is your “risk tolerance”
- Larger R = less risk aversion
  - Makes utility function more “linear”
- \( R \approx \) highest value of Y for which you would play:

```
Play?  yes  0.5  $Y
       no  0.5 -$Y/2
     $0
```
How Rational Are You?

• Which option would you choose?

Choice A preferred:
\[ 1.00 \text{U}(1,000,000) > 0.89 \text{U}(1,000,000) + 0.01 \text{U}(0) + 0.10 \text{U}(5,000,000) \]

Choice D preferred:
\[ 0.89 \text{U}(0) + 0.11 \text{U}(1,000,000) < 0.90 \text{U}(0) + 0.10 \text{U}(5,000,000) \]

• How many chose A and D?

<table>
<thead>
<tr>
<th>Choice</th>
<th>Percentage</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100%</td>
<td>$1,000,000</td>
</tr>
<tr>
<td>B</td>
<td>89%</td>
<td>$1,000,000</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>$5,000,000</td>
</tr>
<tr>
<td>C</td>
<td>89%</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>$1,000,000</td>
</tr>
<tr>
<td>D</td>
<td>90%</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>$5,000,000</td>
</tr>
</tbody>
</table>
How Rational Are You?

• Which option would you choose?

Choice D preferred:
1.00 U(1,000,000) < 0.89 U(1,000,000) + 0.01 U(0) + 0.10 U(5,000,000)

Choice A preferred:
1.00 U(1,000,000) > 0.89 U(1,000,000) + 0.01 U(0) + 0.10 U(5,000,000)

Add 0.89 U(1,000,000) to both sides
Subtract 0.89 U(0) from both sides

X < Y
X > Y

Choice D preferred:
0.11 U(1,000,000) < 0.01 U(0) + 0.10 U(5,000,000)

Choice D preferred:
0.89 U(0) + 0.11 U(1,000,000) < 0.90 U(0) + 0.10 U(5,000,000)

• You are inconsistent with utility theory (Allais Paradox)
  o For any choice of utility function
How Rational Are You?

• Which option would you choose?

A 100% $1,000,000

B 89% $1,000,000
   1% $0

C 89% $0
   11% $1,000,000

D 90% $0
   10% $5,000,000

Choice A preferred: 1.00 U(1,000,000) > 0.89 U(1,000,000) + 0.01 U(0) + 0.10 U(5,000,000)

Choice D preferred: 0.89 U(0) + 0.11 U(1,000,000) < 0.90 U(0) + 0.10 U(5,000,000)

• Human behavior is not always axiomatically consistent
Micromort

- A **micromort** is 1 in 1,000,000 chance of death
  - How much would you need to be paid to take on the risk of a micromort?
  - How much would you pay to avoid a micromort?
    - $P(\text{die in plane crash}) \approx 1 \text{ in } 1,500,000$
    - $P(\text{killed by lightning}) \approx 1 \text{ in } 1,400,000$
  - How much would you need to be paid to take on a decimort (1 in 10 chance of death)?
  - If you think this is morbid, companies actually do this
    - Car manufacturers
    - Insurance companies
Let’s Do a Real Test

• Game set-up
  ▪ I will flip a fair coin
  ▪ If “heads”, you win $50. If “tails”, you win $0
  ▪ How much would you be willing to pay me to play?
    o $1 ?
    o $10 ?
    o $20 ?
    o $24.99 ?
    o $25.01 ?
    o $35 ?
  ▪ Maximal value?
    o Come on down!
    o How did you determine that value?