15: General Inference

Amrita Kaur
July 25, 2022
Let’s talk Midterm

Class Stats
• Median: 103.5/120 (86%)
• Mean: 100/120 (83%)
• Std: 11.5

How to interpret your score:
• Between 100-120 points: Rock on – always a little more to learn :)
• Between 75-99 points: Solid – review those few concepts that you forgot
• Below 74 points: Check-in with us – good time to recalibrate
Resurrection Final

If you end up doing better on the final than you did on the midterm, we will significantly down-weight your midterm grade.

TLDR: we want to reward improvement

*Exception: we won’t give out A+’s using resurrection final grading scheme
Where we left off...
Constructing a Bayesian Network

1. Describe the joint distribution using causality.

2. Provide $P(\text{values}|\text{parents})$ for each random variable.

- $P(F_{ev} = 1|F_{lu} = 0)$
- $P(F_{ev} = 1|F_{lu} = 1)$
- $P(T = 1|F_{lu} = 0, U = 0)$
- $P(T = 1|F_{lu} = 0, U = 1)$
- $P(T = 1|F_{lu} = 1, U = 0)$
- $P(T = 1|F_{lu} = 1, U = 1)$
- $P(F_{lu} = 1)$
- $P(U = 1)$
Bayes Net Assumption

Order nodes by ancestry:

\[ P(\text{Joint}) = \prod_{i} P(x_i|x_{i-1}, \ldots, x_1) \]
\[ = \prod_{i} P(x_i|\text{Values of parents of } X_i) \]
Constructing a Bayesian Network

1. Describe the joint distribution using causality.
2. Provide $P(\text{values}|\text{parents})$ for each random variable.
3. Assume conditional independence
Independent Discrete RVs

Recall the definition of independent events $E$ and $F$:

\[ P(EF) = P(E)P(F) \]
\[ P(E|F) = P(E) \]

Two discrete random variables $X$ and $Y$ are independent if:

for all $x, y$:
\[
P(X = x, Y = y) = P(X = x)P(Y = y)
\]
\[
p_{X,Y}(x, y) = p_X(x)p_Y(y)
\]

Intuitively: knowing value of $X$ tells us nothing about the distribution of $Y$ (and vice versa)
Independence of Multiple RVs

Recall independence of $n$ events $E_1, E_2, \ldots, E_n$:

for $r = 1, \ldots, n$:

for every subset $E_1, E_2, \ldots, E_r$:

\[
P(E_1, E_2, \ldots, E_r) = P(E_1)P(E_2) \cdots P(E_r)
\]

We have independence of $n$ discrete random variables $X_1, X_2, \ldots, X_n$ if for all $x_1, x_2, \ldots, x_n$:

\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)
\]
Conditionally Independent RVs

Recall that two events $A$ and $B$ are conditionally independent given $E$ if:

$$P(AB|E) = P(A|E)P(B|E)$$

$n$ discrete random variables $X_1, X_2, \ldots, X_n$ are called conditionally independent given $Y$ if:

for all $x_1, x_2, \ldots, x_n, y$:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|Y = y) = \prod_{i=1}^{n} P(X_i = x_i|Y = y)$$
Constructing a Bayesian Network

1. Describe the joint distribution using causality.
2. Provide $P(\text{values}|\text{parents})$ for each random variable.
3. Assume conditional independence (When you decide your causality, the independence assumptions come for free.)
How can I make models from data?
<table>
<thead>
<tr>
<th>Music</th>
<th>Dance</th>
<th>Folk</th>
<th>Country</th>
<th>Classical mu</th>
<th>Musical</th>
<th>Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Pop</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Classical mu</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Folk</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Rock</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Jazz</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Latin American</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Stanford University
Spot the difference

Both distributions have the same $E[X]$, $E[Y]$, $\text{Var}(X)$, $\text{Var}(Y)$
Covariance measures how one random variable varies with a second.

- Age and income have a positive covariance.
- Handedness and musical ability have near zero covariance.
- Vehicle age and price have a negative covariance.

The covariance of two variables $X$ and $Y$ is:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$
Properties of Covariance

The covariance of two variables $X$ and $Y$ is:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

$$= E[XY] - E[X]E[Y]$$

Properties:

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(X, X) = E[X^2] - (E[X])^2 = \text{Var}(X)$
3. Covariance of sums = sum of all pairwise covariances
   $$\text{Cov}(X_1 + X_2, Y_1 + Y_2) = \text{Cov}(X_1, Y_1) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_2)$$
4. Covariance is non-linear: $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$
Zero Covariance and Independence

If $X \perp Y$, then $\text{Cov}(X,Y) = 0$

$\text{Cov}(X,Y) = 0$ does not mean that $X \perp Y$
Statistics of sums of RVs

For any random variables $X$ and $Y$,

\[
E[X + Y] = E[X] + E[Y]
\]

\[
\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)
\]

For independent $X$ and $Y$,

\[
E[XY] = E[X]E[Y]
\]

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)
\]
Covarying Humans

What is the covariance of weight $W$ and height $H$?

$$= 3355.83 - (62.75)(52.75)$$
$$= 45.77$$

What about weight (lb) and height (cm)?

$$\text{Cov}(2.20W, 2.54H)$$
$$= E[2.20W \cdot 2.54H] - E[2.20W]E[2.54H]$$
$$= 18752.38 - (138.05)(133.99)$$
$$= 255.06$$
Correlation

The **correlation** of two variables $X$ and $Y$ is:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$\sigma_X^2 = \text{Var}(X), \quad \sigma_Y^2 = \text{Var}(Y)$$

Note: $-1 \leq \rho(X, Y) \leq 1$

Correlation measures the **linear relationship** between $X$ and $Y$:

- $\rho(X, Y) = 1 \implies Y = aX + b$, where $a = \sigma_Y / \sigma_X$
- $\rho(X, Y) = -1 \implies Y = aX + b$, where $a = -\sigma_Y / \sigma_X$
- $\rho(X, Y) = 0 \implies \text{"uncorrelated" (absence of linear relationship)}$
Per capita cheese consumption correlates with Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% (r=0.947091)

Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention
Divorce rate in Maine correlates with Per capita consumption of margarine

Correlation: 99.26% (r=0.992558)

Data sources: National Vital Statistics Reports and U.S. Department of Agriculture
<table>
<thead>
<tr>
<th>Music</th>
<th>Dance</th>
<th>Folk</th>
<th>Country</th>
<th>Classical mu</th>
<th>Musical</th>
<th>Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Rap</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Latin-American</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Jazz</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Folk</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Blues</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Classical</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Folk</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Western</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Western</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Rock</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Country</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Correlation of Music Tastes
Correlation of Music Tastes
Correlation of Music Tastes

[Heatmap diagram showing correlations between various music genres such as metal, rock, pop, punk, and dance.

Stanford University]
How do we solve problems with models?
Inference

An updated belief about a random variable (or multiple) based on conditional knowledge regarding another random variable (or multiple) in a probabilistic model.

Just conditional probability with random variables!
Inference Method #1: Algebra
Inference via math

\[ P(F_{lu} = 1) = 0.1 \quad P(U = 1) = 0.8 \]

1. \[ P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1) \]

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \]
\[ P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]
\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]
Inference via math

\[ P(\text{Flu} = 1) = 0.1 \quad P(U = 1) = 0.8 \]

2. \[ P(\text{Flu} = 1|\text{Fev} = 0, U = 0, T = 1)? \]

\[ P(\text{Fev} = 1|\text{Flu} = 1) = 0.9 \]
\[ P(\text{Fev} = 1|\text{Flu} = 0) = 0.05 \]

\[ P(T = 1|\text{Flu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|\text{Flu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|\text{Flu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|\text{Flu} = 1, U = 1) = 1.0 \]
Inference via math

1. Compute joint probabilities

\[ P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1) \]
\[ P(F_{lu} = 0, F_{ev} = 0, U = 0, T = 1) \]

2. Definition of conditional probability

\[
\frac{P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1)}{\sum_x P(F_{lu} = x, F_{ev} = 0, U = 0, T = 1)} = 0.095
\]
Inference via math

\[ P(\text{Flu} = 1) = 0.1 \]
\[ P(U = 1) = 0.8 \]

3. \[ P(\text{Flu} = 1|U = 1, T = 1) \]?
Inference via math

Infection via math

3. \( P(F_{lu} = 1 | U = 1, T = 1) \)?

1. Compute joint probabilities

\[
P(F_{lu} = 1, U = 1, F_{ev} = 1, T = 1) \quad \ldots \quad P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1)\]

2. Definition of conditional probability

\[
\frac{\sum_y P(F_{lu} = 1, U = 1, F_{ev} = y, T = 1)}{\sum_x \sum_y P(F_{lu} = x, U = 1, F_{ev} = y, T = 1)} = 0.122
\]
Inference Method #2: Rejection Sampling

https://colab.research.google.com/drive/1ijqOVWWAwSWIrIPGpjHo6r6HJ6L1HYXp?usp=sharing