Midterm Location by Last Name

CEMEX: A through K
Hewlett 200: L through S
Hewlett 201: T through Z

Look out for your assignment! Let us know if you are at risk
No Class on Monday

This is how I imagine students are when they are studying 😊

(recall why we are here. Learning is a joy. You are growing)
Computers Couldn’t Understand Student Code

60,000 students attempted this problem
37,000 unique solutions

Challenge

Student Code

Insight

You need to move and turn in your loop
Computers Couldn’t Understand Student Code

![Diagram showing feedback F1 score for different models and humans. The chart compares Cond. Prob., Deep Learning, Generative Model, and Humans in terms of their ability to understand Code.org short code (10 lines).]
Generative Model of Characters

Lake et al, 2015

procedure GENERATETYPE
    $\kappa \leftarrow P(\kappa)$ \> Sample number of parts
    for $i = 1 \ldots \kappa$ do
        $n_i \leftarrow P(n_i | \kappa)$ \> Sample number of sub-parts
        for $j = 1 \ldots n_i$ do
            $s_{ij} \leftarrow P(s_{ij} | s_{i(j-1)})$ \> Sample sub-part sequence
        end for
        $R_i \leftarrow P(R_i | S_1, \ldots, S_{i-1})$ \> Sample relation
    end for
    $\psi \leftarrow \{\kappa, R, S\}$
    return @GENERATE_TOKEN($\psi$) \> Return program

procedure GENERATE_TOKEN($\psi$)
    for $i = 1 \ldots \kappa$ do
        $S_i^{(m)} \leftarrow P(S_i^{(m)} | S_i)$ \> Add motor variance
        $L_i^{(m)} \leftarrow P(L_i^{(m)} | R_i, T_i^{(m)}, \ldots, T_{i-1}^{(m)})$ \> Sample part's start location
        $T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)})$ \> Compose a part's trajectory
    end for
    $A^{(m)} \leftarrow P(A^{(m)})$ \> Sample affine transform
    $f^{(m)} \leftarrow P(f^{(m)} | T^{(m)}, A^{(m)})$ \> Sample image
    return $f^{(m)}$
Computers Couldn’t Understand Student Code

Code.org Short Code (10 lines)

Feedback F1 Score

Cond. Prob.  Deep Learning  Generative Model  Humans

Outstanding Student paper award, AAAI 2019
Review
Last Week: Joint Distributions

Joint Distribution noun

The probability of a simultaneous assignment to all the random variables in a probabilistic model.

Eg:

\[ P(X = x, Y = y) \]
\[ f(X = x, Y = y) \]
\[ P(X = x, Y = y, \cdots, Z = z) \]
Inference noun

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

TLDR: conditional probability with random variables.
Inference with Continuous

Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?
Does this define the joint?

\[ f(G = g, X = x) = f(X = x|G = g)P(G = g) \]

Q: What is \( P(G = 1 \mid X = 163) \)
Harder when inferred random variable is not a Bernoulli
Inference on a non-bernoulli random variable

\[ P(A = a) \]

Observation \( Y = 0 \)

\[ P(A = a|E) \]

Known function. Also needs font size (constant)

\[
P(A = a|Y = 0) = \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)}
\]
Number or Dictionary?

\[
P(A = a | Y = 0) = \frac{P(Y = 0 | A = a)P(A = a)}{P(Y = 0)}
\]

belief[a] = 0.001
Inference on a non-bernoulli random variable

In plain English: run bayes for each value of $a$

$$P(A = a | Y = 0) = \frac{P(Y = 0 | A = a)P(A = a)}{P(Y = 0)}$$

# RV bayes as code

def update(belief, obs):
    for a in support:
        prior_a = belief[a]
        likelihood = calc_likelihood(a, obs)
        belief[a] = prior_a * likelihood
        normalize(belief)
# RV bayes as code

```python
def update(belief, obs):
    for a in support:
        prior_a = belief[a]
        likelihood = calc_likelihood(a, obs)
        belief[a] = prior_a * likelihood
    normalize(belief)
```

In plain English: this is the numerator, summed over all values of A.

\[
P(A = a|Y = 0) = \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)}
\]

\[
= \frac{P(Y = 0|A = a)P(A = a)}{\sum_{x \in A} P(Y = 0, A = x)}
\]

\[
= \frac{P(Y = 0|A = a)P(A = a)}{\sum_{x \in A} P(Y = 0|A = x)P(A = x)}
\]
1. Perspective on the artform of how to design probabilistic models
2. How to calculate Correlations
3. Use and verify Independence with Random Variables
Let's talk about how to make a model
Model Version #1: Python That Outputs a **Joint** Sample

![Python Logo]

Sample Baby Elephant

- Sex: Female
- Weight: 161kg
Model Version #2: Bayesian Network

Does this define the joint?

\[ f(G = g, X = x) = f(X = x | G = g) P(G = g) \]

\( G = 1 \) is Bern(p = 0.5)

\( X | G = 1 \) is N(\( \mu = 160 \), \( \sigma^2 = 7^2 \))

\( X | G = 0 \) is N(\( \mu = 165 \), \( \sigma^2 = 3^2 \))
Why You Need a Model
What are your symptoms?

Type your main symptom here
Inference

**Inference question:**
Given the values of some random variables, what are the conditional distributions of some other random variables?
Inference

One inference question:

\[ P(F = 1|N = 1, T = 1) = \frac{P(F = 1, N = 1, T = 1)}{P(N = 1, T = 1)} \]
Inference

Another inference question:

\[ P(C_o = 1, U = 1 | S = 0, F_e = 0) \]

\[ = \frac{P(C_o = 1, U = 1, S = 0, F_e = 0)}{P(S = 0, F_e = 0)} \]
Inference

If we knew the joint distribution, we can answer all probabilistic inference questions.

What is the size of the joint probability table?

- A. $2^{N-1}$ entries
- B. $N^2$ entries
- C. $2^N$ entries
- D. None/other/don’t know

$N = 9$

all binary RVs
Inference

If we knew the joint distribution, we can answer all probabilistic inference questions.

What is the size of the joint probability table?

A. $2^{N-1}$ entries
B. $N^2$ entries
C. $2^N$ entries
D. None/other/don’t know

Naively specifying a joint distribution is, in general, intractable.
N can be large...
Bayesian Networks
A simpler WebMD

Great! Just specify $2^4 = 16$ joint probabilities...?

$P(F_{lu} = a, F_{ev} = b, U = c, T = d)$

What would a Stanford flu expert do?

Describe the joint distribution using causality!
What would a Stanford flu expert do?

1. Describe the joint distribution using causality.
2. Provide $P(\text{values}|\text{parents})$ for each random variable
3. Implicitly assumes independences.
Constructing a Bayesian Network

In a Bayesian Network,

Each random variable is caused by its parents. Def $P(\text{node} \mid \text{parents})$

- Node: random variable
- Directed edge: causality

Examples:

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

$P(T = 1|F_{lu} = 0, U = 0)$
$P(T = 1|F_{lu} = 0, U = 1)$
$P(T = 1|F_{lu} = 1, U = 0)$
$P(T = 1|F_{lu} = 1, U = 1)$
Constructing a Bayesian Network

What would a Stanford flu expert do?

✅ 1. Describe the joint distribution using causality.

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

✅ 3. Implicitly assumes independences.

![Bayesian Network Diagram]

- Flu
- Under-grad
- Fever
- Tired
This model assumes that Flu and being an Undergraduate are independent.

**Neat trick:** it also assumes that fever and tired are conditionally independent given Flu.

You need to tell a generative story. You do **not** need to be able to reason about all the implied independencies.
Bug: Constructing a Bayesian Network

Flu → Under-grad → Tired → Fever

Must by acyclic!
Model Version #1: Python That Outputs a **Joint** Sample

Sample Baby Elephant
Sex: Female
Weight: 161kg
Bayesian Network Assumption:

\[ P(\text{Joint}) = P(X_1 = x_1 \ldots X_n = x_n) = \prod_{i} P(X_i = x_i | \text{parents of } X_i) \]
Bayes Nets tell a generative story.

This leads to many independence assumptions

Makes it **tractable** to represent the joint
Challenge: Exact Inference in a Bayes Net

\[ P(Fl = 0 | Fe = 1) \]

\[ = \frac{P(Fl = 0, Fe = 1)}{P(Fe = 1)} \]

\[ = \frac{P(Fe = 1 | Fl = 0) P(Fl = 0)}{\sum_i P(Fe = 1 | Fl = i) P(Fl = i)} \]
Independence of RVs
Independent discrete RVs

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

$P(EF) = P(E)P(F)$

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

for all $x, y$:

$P(X = x, Y = y) = P(X = x)P(Y = y)$

$p(x, y) = p(x)p(y)$

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

• If two variables are not independent, they are called dependent.
Dice (after all this time, still our friends)

Let: $D_1$ and $D_2$ be the outcomes of two rolls $S = D_1 + D_2$, the sum of two rolls

- Each roll of a fair, 6-sided die is an independent trial.
- Random variables $D_1$ and $D_2$ are independent.

1. Are events $(D_1 = 1)$ and $(S = 7)$ independent?

2. Are events $(D_1 = 1)$ and $(S = 5)$ independent?

3. Are random variables $D_1$ and $S$ independent?
Dice (after all this time, still our friends)

Let: $D_1$ and $D_2$ be the outcomes of two rolls, $S = D_1 + D_2$, the sum of two rolls

- Each roll of a 6-sided die is an independent trial.
- Random variables $D_1$ and $D_2$ are independent.

1. Are events $(D_1 = 1)$ and $(S = 7)$ independent? ✓
2. Are events $(D_1 = 1)$ and $(S = 5)$ independent? ✗
3. Are random variables $D_1$ and $S$ independent? ✗

All events $(X = x, Y = y)$ must be independent for $X, Y$ to be independent RVs.
Can I discover independence from data?
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>Dance</td>
<td>Folk</td>
<td>Country</td>
<td>Classical music</td>
<td>Musical</td>
<td>Pop</td>
<td>Rock</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>27</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>28</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>29</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>31</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>33</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>34</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
Why is it harder to find independences here than for bat DNA expression?
Dance of the Covariance
Recall our Ebola Bats
## Bat Data

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
<th>Gene3</th>
<th>Gene4</th>
<th>Gene5</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

...
## Expression Amount

<table>
<thead>
<tr>
<th>Gene5</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.76</td>
<td>0.83</td>
</tr>
<tr>
<td>0.94</td>
<td>0.85</td>
</tr>
<tr>
<td>0.82</td>
<td>0.03</td>
</tr>
<tr>
<td>0.94</td>
<td>0.32</td>
</tr>
<tr>
<td>0.50</td>
<td>0.10</td>
</tr>
<tr>
<td>0.40</td>
<td>0.53</td>
</tr>
<tr>
<td>0.90</td>
<td>0.67</td>
</tr>
<tr>
<td>0.29</td>
<td>0.71</td>
</tr>
<tr>
<td>0.72</td>
<td>0.25</td>
</tr>
<tr>
<td>0.15</td>
<td>0.24</td>
</tr>
<tr>
<td>0.79</td>
<td>0.98</td>
</tr>
<tr>
<td>0.68</td>
<td>0.77</td>
</tr>
<tr>
<td>0.71</td>
<td>0.37</td>
</tr>
<tr>
<td>0.36</td>
<td>0.18</td>
</tr>
<tr>
<td>0.62</td>
<td>0.08</td>
</tr>
<tr>
<td>0.59</td>
<td>0.38</td>
</tr>
<tr>
<td>0.82</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Vary Together

\[ x - E[x] = 3 \]

\[ y - E[y] = 2.6 \]

\[ (x - E[x])(y - E[y]) = 7.8 \]
Vary Together

\[ (x - E[x])(y - E[y]) = 0 \]

\[ x - E[x] \approx 0 \]

\[ y - E[y] \approx 0 \]
Vary Together

\[ x - E[x] = -1.1 \]
\[ y - E[y] = -2.8 \]

\[(x - E[x])(y - E[y]) \approx 3.1\]
Understanding Covariance
Say X and Y are arbitrary random variables

Covariance of X and Y:

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

The Dance of the Covariance

Above mean Above mean Positive

Bellow mean Above mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Above mean Above mean Positive

Bellow mean Bellow mean Positive

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Above mean Positive

Bellow mean Bellow mean Positive

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Bellow mean Above mean Negative

Above mean Bellow mean Negative

Below mean Above mean Positive

Bellow mean Bellow mean Positive

Bellow mean Above mean Negative

Above mean Bellow mean Negative
Say X and Y are arbitrary random variables

Covariance of X and Y:

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

Equivalently:


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

- X and Y independent, $E[XY] = E[X]E[Y] \Rightarrow \text{Cov}(X, Y) = 0$
- But $\text{Cov}(X, Y) = 0$ does not imply X and Y independent!
Consider the following data:

<table>
<thead>
<tr>
<th>Weight</th>
<th>Height</th>
<th>Weight * Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

\[
E[W] = 62.75 \\
E[H] = 52.75 \\
E[W*H] = 3355.83
\]

\[
= 3355.83 - (62.75)(52.75) \\
= 45.77
\]
Covariance

Poll: (a) positive, (b) negative, (c) zero
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero

Positive
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero

Negative
Is the Covariance: (a) positive, (b) negative, (c) zero
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero
X and Y are random variables with PMF:

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
<th>p_X(x)</th>
<th>p_Y(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-1</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
</tr>
<tr>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
</tbody>
</table>

\[
Y = \begin{cases} 
0 & \text{if } X \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]

- \( E[X] = -1(1/3) + 0(1/3) + 1(1/3) = 0 \)
- \( E[Y] = 0(2/3) + 1(1/3) = 1/3 \)
- Since \( XY = 0 \), \( E[XY] = 0 \)
- \( \text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 = 0 \)

But, X and Y are clearly dependent!
Properties of Covariance

Say X and Y are arbitrary random variables

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = E[X^2] - E[X]E[X] = \text{Var}(X)$
- $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$
Correlation
Consider the following data:

<table>
<thead>
<tr>
<th>Weight</th>
<th>Height</th>
<th>Weight * Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

\[
\text{Cov}(W, H) = \mathbb{E}[W\times H] - \mathbb{E}[W]\mathbb{E}[H]
\]
\[
= 3355.83 - (62.75)(52.75)
\]
\[
= 45.77
\]
Cauchy Schwarz, a great way to normalize!

\[-\text{Std}(X)\text{Std}(Y) \leq \text{Cov}(X, Y) \leq \text{Std}(X)\text{Std}(Y)\]
Say $X$ and $Y$ are arbitrary random variables

- Correlation of $X$ and $Y$, denoted $\rho(X, Y)$:
  
  $$
  \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}
  $$

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

- Correlation measures **linearity** between $X$ and $Y$
  
  - $\rho(X, Y) = 1 \implies Y = aX + b$ where $a = \frac{\sigma_y}{\sigma_x}$
  - $\rho(X, Y) = -1 \implies Y = aX + b$ where $a = -\frac{\sigma_y}{\sigma_x}$
  - $\rho(X, Y) = 0 \implies$ absence of **linear** relationship

- **But**, $X$ and $Y$ can still be related in some other way!

- **If** $\rho(X, Y) = 0$, we say $X$ and $Y$ are “uncorrelated”
  
  - **Note**: Independence implies uncorrelated, but **not** vice versa!
Say X and Y are arbitrary random variables

- Correlation of X and Y, denoted $\rho(X, Y)$:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Say $Y = cX$. Correlation should be 1.
Correlation of Music Tastes
Correlation of Music Tastes
From Correlation to Bayes Net. Alternative!
From Correlation to Bayes Net. Alternative!

music
  ┌───┐
  │   │
  │ classy │ dancy │ rocky │ funky │ folky │
  │   │
  └───┘

categories
  ┌───┐
  │   │
  │ opera │ punk │ reggae │ country │
  │   │
  └───┘

Chris Piech, CS109, 2021
From Correlation to Bayes Net. Alternative!
From Correlation to Bayes Net. Alternative!
How do you know if your model is good?

Answer: it is accurate at inference (especially tasks you care about)
Rock Music Vs Oil?

High Correlation

Hubbert Peak Theory

http://www.aei.org/publication/blog/
Tell your friends!

![Graph showing correlation between per capita cheese consumption and number of deaths due to entanglement in bedsheets](image)

<table>
<thead>
<tr>
<th>Year</th>
<th>Per capita cheese consumption (US)</th>
<th>Number of deaths due to entanglement in bedsheets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>29.8 Pounds</td>
<td>327 Deaths</td>
</tr>
<tr>
<td>2001</td>
<td>30.1 Pounds</td>
<td>456 Deaths</td>
</tr>
<tr>
<td>2002</td>
<td>30.5 Pounds</td>
<td>509 Deaths</td>
</tr>
<tr>
<td>2003</td>
<td>30.6 Pounds</td>
<td>497 Deaths</td>
</tr>
<tr>
<td>2004</td>
<td>31.3 Pounds</td>
<td>596 Deaths</td>
</tr>
<tr>
<td>2005</td>
<td>31.7 Pounds</td>
<td>573 Deaths</td>
</tr>
<tr>
<td>2006</td>
<td>32.6 Pounds</td>
<td>661 Deaths</td>
</tr>
<tr>
<td>2007</td>
<td>33.1 Pounds</td>
<td>741 Deaths</td>
</tr>
<tr>
<td>2008</td>
<td>32.7 Pounds</td>
<td>809 Deaths</td>
</tr>
<tr>
<td>2009</td>
<td>32.8 Pounds</td>
<td>717 Deaths</td>
</tr>
</tbody>
</table>

Correlation: 0.947091

Sources: USDA & CDC, tylervigen.com
What about Code.org?
Computers Couldn’t Understand Code

60,000 students attempted this problem
37,000 unique solutions

Challenge

Student Code

You need to move and turn in your loop

Insight
Computers Couldn’t Understand Code

Code.org Short Code (10 lines)

Feedback F1 Score

Humans

Generative Model

Deep Learning

Cond. Prob.
Code is hard to think about...

**Code Zipf Plot**

\[ f(k) = \frac{1/k^s}{\sum_n^N (1/n^s)} \]

*Exponential combination of decisions. Super fat tailed. Everything looks unique.*
Generative Model is Easier than Straight Inference

This is easy and exponential

\[ \Theta \sim \text{pythonSample} \]

\[ C \sim \text{pythonSample}|\Theta \]

\[ \Pi \sim \text{pythonSample}|C \]

This is hard and linear

\[ P(\Theta, C|\Pi) \]

Infer ability and choices from code
Made a probabilistic programming language:

ideaToText
Write a model of each **decision point**, given parents

1. This is code for a single decision point

```python
# This python class is a RubricSampling Decision
# it generates programs that print the numbers 10 -> 1
class Countdown(Decision):
    def registerChoices(self):
        # these are the main strategies for printing out a
countdown
        self.addRubricChoices('loop-style', {
            'for': 100,
            'while': 40,
            'none': 1,
            'empty': 1
        })

    def processChoices(self):
        # we can make some grading choices based on which
        # strategy they chose (did they actually use a loop?)
        style = self.getChoice('loop-style')
        hasLoop = style != 'none' and style != 'empty'
        self.addLabel('rubric-hasLoop', hasLoop)
        self.addLabel('rubric-printsNums', style != 'empty')

    def renderCode(self):
        style = self.getChoice('loop-style')
        if style == 'for': return '{ForSoln}
        if style == 'while': return '{WhileSoln}
        if style == 'none': return '{NoLoopSoln}
        if style == 'empty': return ''
```

2. Give a name to the choice that the student is making

3. How do those choices translate into grades?

4. What does the code look like? Often evokes other decision points
Generative Model of Grading

Lake et al, 2015

```
procedure GENERATETYPE
    \( \kappa \leftarrow P(\kappa) \) \hspace{1cm} \text{Sample number of parts}
    \( n_i \leftarrow P(n|\kappa) \) \hspace{1cm} \text{Sample number of sub-parts}
    \( \kappa_j \leftarrow P(\kappa|\kappa_{j-1}) \) \hspace{1cm} \text{Sample sub-part sequence}
    \( R_i \leftarrow P(R_i|S_1, \ldots, S_{i-1}) \) \hspace{1cm} \text{Sample relation}
    \psi \leftarrow \{s, R, S\} \hspace{1cm} \text{return @GENERATE_TOKEN(\psi)} \hspace{1cm} \text{Return program}
```

Muke Wu, Ali Malik, Noah Goodman, Chris Piech, 2019

```
procedure GENERATE_TOKEN(\psi)
    \( x^{(0)} \leftarrow P(s^{(0)}|S) \) \hspace{1cm} \text{Add motor variance}
    \( l^{(0)} \leftarrow P(l^{(0)}|R_0, T^{(0)}) \) \hspace{1cm} \text{Sample part's start location}
    \( z^{(0)} \leftarrow f(l^{(0)}, s^{(0)}) \) \hspace{1cm} \text{Compose a part's trajectory}
    \( A^{(0)} \leftarrow P(A^{(0)}) \) \hspace{1cm} \text{Sample affine transform}
    \( f^{(0)} \leftarrow P(f^{(0)}|T^{(0)}, A^{(0)}) \) \hspace{1cm} \text{Sample image}
    \text{return } f^{(0)}
```

Outstanding Student paper award, AAAI 2019
Computers Couldn’t Understand Code

Code.org Short Code (10 lines)

- Cond. Prob.
- Deep Learning
- Generative Model
- Humans
Can A.I. Grade Your Next Test?

Neural networks could give online education a boost by providing automated feedback to students.

Friday: General Inference

Generative model

Decision process

Inference model

Output solution
Que te vayas bien
No Class on Monday

This is how I imagine students are when they are studying 😊

(recall why we are here. Learning is a joy. You are growing)