Netflix intro

What data should we encode about each Netflix account holder to help us make effective recommendations?

In machine learning, clustering can be used to group similar data for prediction and recommendation. For example, each Netflix user's viewing history can be represented as a n-tuple indicating their preferences about movies in the database, where n is the number of movies in the database. People with similar tastes in movies can then be clustered to provide recommendations of movies for one another. Mathematically, clustering is based on a notion of distance between pairs of n-tuples.

Infinite finite sets def

 \mathbb{Z} The set of integers

 \mathbb{Z}^+ The set of positive integers

The set of nonnegative integers \mathbb{N}

The set of rational numbers \mathbb{Q}

 \mathbb{R} The set of real numbers

$$\left\{\dots, -2, -1, 0, 1, 2, \dots\right\}$$

$$\left\{1, 2, \dots\right\}$$

$$\left\{0, 1, 2, \dots\right\}$$

$$\left\{\frac{p}{q} \mid p \in \mathbb{Z} \text{ and } q \in \mathbb{Z} \text{ and } q \neq 0\right\}$$

The above sets are all **infinite**.

A finite set is one whose distinct elements can be counted by a natural number.

Examples of finite sets: \emptyset , $\{\sqrt{2}\}$

Motivating question: Are some of the above sets bigger than others?

Musical chairs analogy

Analogy: Musical chairs



People try to sit down when the music stops

Person♥ sits in Chair 1, Person® sits in Chair 2,

Person© is left standing!

What does this say about the number of chairs and the number of people?

Defined functions

Defining functions A function is defined by its (1) domain, (2) codomain, and (3) rule assigning each element in the domain exactly one element in the codomain. The domain and codomain are nonempty sets. The rule can be depicted as a table, formula, English description, etc.

(Rosen p139)

Example: $f_A: \mathbb{R}^+ \to \mathbb{Q}$ with $f_A(x) = x$ is **not** a well-defined function because

Example: $f_B: \mathbb{Q} \to \mathbb{Z}$ with $f_B\left(\frac{p}{q}\right) = p + q$ is **not** a well-defined function because

Example: $f_C: \mathbb{Z} \to \mathbb{R}$ with $f_C(x) = \frac{x}{|x|}$ is **not** a well-defined function because

Injective cardinality definition

Definition (Rosen p141): A function $f: D \to C$ is **one-to-one** (or injective) means for every a, b in the domain D, if f(a) = f(b) then a = b.

Definition: For sets A, B, we say that the cardinality of A is no bigger than the cardinality of B, and write $|A| \leq |B|$, to mean there is a one-to-one function with domain A and codomain B.

Injective cardinality musical chairs

In the analogy: The function $sitter: \{Chair1, Chair2\} \rightarrow \{Person \heartsuit, Person \heartsuit, Person \heartsuit\}$ given by $sitter(Chair1) = Person \heartsuit, sitter(Chair2) = Person \heartsuit$, is one-to-one and witnesses that

 $|\{Chair1, Chair2\}| \leq |\{Person \heartsuit, Person \heartsuit, Person \heartsuit\}|$

Rna injective cardinality

Let S_2 be the set of RNA strands of length 2.

Statement	True/False , justification
$ \{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\} \leq S_2 $	
$ \{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\}\times\{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\} \leq S_2 $	

Surjective cardinality definition

Definition (Rosen p143): A function $f: D \to C$ is **onto** (or surjective) means for every b in the codomain, there is an element a in the domain with f(a) = b.

Formally, $f: D \to C$ is onto means _____.

Definition: For sets A, B, we say that the cardinality of A is no smaller than the cardinality of B, and write $|A| \ge |B|$, to mean there is an onto function with domain A and codomain B.

Surjective cardinality musical chairs

In the analogy: The function $triedToSit: \{Person \heartsuit, Person \heartsuit, Person \heartsuit\} \rightarrow \{Chair1, Chair2\}$ given by $triedToSit(Person \heartsuit) = Chair1, triedToSit(Person \heartsuit) = Chair2, triedToSit(Person \heartsuit) = Chair2,$ is onto and witnesses that

 $|\{Person \circlearrowleft, Person \odot, Person \odot\}| \geq |\{Chair1, Chair2\}|$

Rna surjective cardinality

Let S_2 be the set of RNA strands of length 2.

Statement	True/False, justification
$ S_2 \geq \{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\} $	
$ S_2 \ge \{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\} \times \{\mathtt{A},\mathtt{U},\mathtt{G},\mathtt{C}\} $	

Bijection definition

Definition (Rosen p144): A function $f: D \to C$ is a **bijection** means that it is both one-to-one and onto. The **inverse** of a bijection $f: D \to C$ is the function $g: C \to D$ such that g(b) = a iff f(a) = b.

For nonempty sets A,B we say $|A| \leq |B| \text{ means there is a one-to-one function with domain } A, \text{ codomain } B$ $|A| \geq |B| \text{ means there is an onto function with domain } A, \text{ codomain } B$ |A| = |B| means there is a bijection with domain A, codomain B

Cardinality properties

Properties of cardinality

$$\begin{split} &\forall A\ (\ |A|=|A|\)\\ &\forall A\ \forall B\ (\ |A|=|B|\ \rightarrow\ |B|=|A|\)\\ &\forall A\ \forall B\ \forall C\ (\ (|A|=|B|\ \land\ |B|=|C|)\ \rightarrow\ |A|=|C|\) \end{split}$$

Extra practice with proofs: Use the definitions of bijections to prove these properties.

Cantor schroder bernstein theorem

Cantor-Schroder-Bernstein Theorem: For all nonempty sets,

|A| = |B| if and only if $(|A| \le |B| \text{ and } |B| \le |A|)$ if and only if $(|A| \ge |B| \text{ and } |B| \ge |A|)$

To prove |A| = |B|, we can do any **one** of the following

- Prove there exists a bijection $f: A \to B$;
- Prove there exists a bijection $f: B \to A$;
- Prove there exists two functions $f_1: A \to B$, $f_2: B \to A$ where each of f_1, f_2 is one-to-one.
- Prove there exists two functions $f_1: A \to B$, $f_2: B \to A$ where each of f_1, f_2 is onto.

Logical equivalence

Logical equivalence Two compound propositions are logically equivalent

means that they have the same truth values for all set-

tings of truth values to their propositional variables.

Tautology A compound proposition that evaluates to true for all

settings of truth values to its propositional variables; it

is abbreviated T.

Contradiction A compound proposition that evaluates to false for all

settings of truth values to its propositional variables; it

is abbreviated F.

Contingency A compound proposition that is neither a tautology nor

a contradiction.

Logical equivalence extra note

Can replace p and q with any compound proposition

Consistency def

Definition: A collection of compound propositions is called **consistent** if there is an assignment of truth values to the propositional variables that makes each of the compound propositions true.

Consistency example

Consistency:

Whenever the system software is being upgraded, users cannot access the file system. If users can access the file system, then they can save new files. If users cannot save new files, then the system software is not being upgraded.

- 1. Translate to symbolic compound propositions
- 2. Look for some truth assignment to the propositional variables for which all the compound propositions output T

Redundancy algorithm

Consider the following algorithm to introduce redundancy in a string of 0s and 1s.

Create redundancy by repeating each bit three times

```
procedure redun3(a_{k-1}\cdots a_0): a binary string)

for i:=0 to k-1

c_{3i}:=a_i

c_{3i+1}:=a_i

c_{3i+2}:=a_i

return c_{3k-1}\cdots c_0
```

Decode sequence of bits using majority rule on consecutive three bit sequences

```
procedure decode3(c_{3k-1}\cdots c_0): a binary string whose length is an integer multiple of 3)

for i:=0 to k-1

if exactly two or three of c_{3i}, c_{3i+1}, c_{3i+2} are set to 1

a_i:=1

else

a_i:=0

return a_{k-1}\cdots a_0
```

Give a recursive definition of the set of outputs of the redun3 procedure, Out,

Basis step: ______

Recursive step: ______

Consider the message m = 0001 so that the sender calculates redun3(m) = redun3(0001) = 000000000111.

Introduce ____ errors into the message so that the signal received by the receiver is _____ but the receiver is still able to decode the original message.

Challenge: what is the biggest number of errors you can introduce?

Building a circuit for line 3 in *decode* procedure: given three input bits, we need to determine whether the majority is a 0 or a 1.

Circuit

c_{3i}	c_{3i+1}	c_{3i+2}	a_i
1	1	1	
1	1	0	
1	0	1	
1	0	0	
0	1	1	
0	1	0	
0	0	1	
0	0	0	

Predicate definition

Definition: A **predicate** is a function from a given set (domain) to $\{T, F\}$.

A predicate can be applied, or evaluated at, an element of the domain.

Predicate equivalent definition

Two predicates over the same domain are **equivalent** means they evaluate to the same truth values for all possible assignments of domain elements to the input.

Predicate truth tables

Input	Output		
	P(x)	N(x)	Mystery(x)
x	$[x]_{2c,3} > 0$	$[x]_{2c,3} < 0$	
000	F		T
001	T		T
010	T		T
011	T		F
100	F		F
101	F		T
110	F		F
111	F		T

The domain for each of the predicates P(x), N(x), Mystery(x) is ______.

Truth set definition

Definition : The truth set of a predicate is the collec	ction of all elements in its domain where the pr	redicate
evaluates to T .		

		•
Truth	COT	exercise
TI UUII	SCU	CACICISC

The truth set for the predicate $P(x)$ is	<u>.</u>
The truth set for the predicate $N(x)$ is	
The truth set for the predicate $Mystery(x)$ is	

Quantification definition

Definitions	(Rosen	40-45):

Example:

The universal quantification of P(x) is the statement "P(x) for all values of x in the domain" and is written $\forall x P(x)$. An element for which P(x) = F is called a **counterexample** of $\forall x P(x)$.

The existential quantification of P(x) is the statement "There exists an element x in the domain such that P(x)" and is written $\exists x P(x)$. An element for which P(x) = T is called a witness of $\exists x P(x)$.

Example:	is a true existential quantification
----------	--------------------------------------

Quantification logical equivalence

Statements involving predicates and quantifiers are logically equivalent means they have the same truth value no matter which predicates (domains and functions) are substituted in.

Quantifier version of De Morgan's laws: $|\neg \forall x P(x) \equiv \exists x (\neg P(x))|$ $\neg \exists x Q(x) \equiv \forall x (\neg Q(x))$ is a false universal quantification. It is logically equivalent to

Rna strand recall

Recall: Each RNA strand is a string whose symbols are elements of the set $B = \{A, C, G, U\}$. The **set of all RNA strands** is called S. The function *rnalen* that computes the length of RNA strands in S is:

Basis Step: If $b \in B$ then $rnalen(S) \rightarrow \mathbb{Z}^+$

Recursive Step: If $s \in S$ and $b \in B$, then rnalen(sb) = 1 + rnalen(s)

Predicate rna example

Example predicates on S

H(s) = T	Truth set of H is
$L_3(s) = \begin{cases} T & \text{if } rnalen(s) = 3\\ F & \text{otherwise} \end{cases}$	Strand where L_3 evaluates to T is e.g
	Strand where L_3 evaluates to F is e.g
$F_{\mathtt{A}}$ is defined recursively by: Basis step: $F_{\mathtt{A}}(\mathtt{A}) = T, \ F_{\mathtt{A}}(\mathtt{C}) = F_{\mathtt{A}}(\mathtt{G}) = F_{\mathtt{A}}(\mathtt{U}) = F$ Recursive step: If $s \in S$ and $b \in B$, then $F_{\mathtt{A}}(sb) = F_{\mathtt{A}}(s)$	Strand where $F_{\mathtt{A}}$ evaluates to T is e.g Strand where $F_{\mathtt{A}}$ evaluates to F is e.g
$P_{\texttt{AUC}}$ is defined as the predicate whose truth set is the collection of RNA strands where the string AUC is a substring (appears inside s , in order and consecutively)	Strand where P_{AUC} evaluates to T is e.g Strand where P_{AUC} evaluates to F is e.g

Cartesian product definition

Definition (Rosen p123): The **Cartesian product** of the sets A and B, $A \times B$, is the set of all ordered pairs (a,b), where $a \in A$ and $b \in B$. That is: $A \times B = \{(a,b) \mid (a \in A) \land (b \in B)\}$. The Cartesian product of the sets A_1, A_2, \ldots, A_n , denoted by $A_1 \times A_2 \times \cdots \times A_n$, is the set of ordered n-tuples (a_1, a_2, \ldots, a_n) , where a_i belongs to A_i for $i = 1, 2, \ldots, n$. That is, $A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i \text{ for } i = 1, 2, \ldots, n\}$

Rna strand recall

Recall: Each RNA strand is a string whose symbols are elements of the set $B = \{A, C, G, U\}$. The **set of all RNA strands** is called S. The function *rnalen* that computes the length of RNA strands in S is:

Basis Step: If $b \in B$ then $rnalen: S \to \mathbb{Z}^+$ rnalen(b) = 1

Recursive Step: If $s \in S$ and $b \in B$, then rnalen(sb) = 1 + rnalen(s)

Rna basecount example

A function basecount that computes the number of a given base b appearing in a RNA strand s is:

 $basecount: S \times B \rightarrow \mathbb{N}$ Basis Step: If $b_1 \in B, b_2 \in B$ $basecount(b_1, b_2) = \begin{cases} 1 & \text{when } b_1 = b_2 \\ 0 & \text{when } b_1 \neq b_2 \end{cases}$ Recursive Step: If $s \in S, b_1 \in B, b_2 \in B$ $basecount(sb_1, b_2) = \begin{cases} 1 + basecount(s, b_2) & \text{when } b_1 = b_2 \\ basecount(s, b_2) & \text{when } b_1 \neq b_2 \end{cases}$

L with domain $S \times \mathbb{Z}^+$ is defined by, for $s \in S$ and $n \in \mathbb{Z}^+$,

BC with domain _____ is defined by, for $s \in S$ and $b \in B$ and $n \in \mathbb{N}$,

$$L(s,n) = \begin{cases} T & \text{if } rnalen(s) = n \\ F & \text{otherwise} \end{cases}$$

$$BC(s, b, n) = \begin{cases} T & \text{if } basecount(s, b) = n \\ F & \text{otherwise} \end{cases}$$

Element where L evaluates to T:

Element where BC evaluates to T:

Element where L evaluates to F:

Element where BC evaluates to F:

Predicate notation

Notation: for a predicate P with domain $X_1 \times \cdots \times X_n$ and a n-tuple (x_1, \ldots, x_n) with each $x_i \in X$, we write $P(x_1, \ldots, x_n)$ to mean $P((x_1, \ldots, x_n))$.

Rna basecount example two

$\exists t \ BC(t)$	In English: _	
Witness that	proves this exi	stential quantification is true:
$\forall (s,b,n) \ (\ BC$	C(s,b,n))	In English:
Counterexample that proves this universal quantification is false:		
New predicates from old $BC(s, b, n)$ means $basecount(s, b) = n$.		

Predicate	Domain	Example domain element where predicate is T
basecount(s,b) = 3		
$basecount(s, \mathbf{A}) = n$		
$\exists n \in \mathbb{N} \ (basecount(s,b) = n)$		
$\forall b \in B \ (basecount(s,b) = 1)$		

Alternating quantifiers

Alternating quar	ntifiers	$\forall s \exists n \ BC(s, \mathtt{A}, n)$		
In English:				
		$\exists n \forall s \ BC(s, \mathtt{U}, n)$		
In English:				
Evaluate each quar	ntified statement as T	or F .		ı
	$\forall s \ \forall b \ \exists n \ BC(s,b,n)$			
	$\exists s \ \forall b \ \exists n \ BC(s,b,n)$	$\forall s \; \exists n \; \forall b \; BC(s,b,n)$	$\exists b \ \exists n \ \forall s \ BC(s,b,n)$	
Extra example: Write the negation of each of the statements above, and use De Morgan's law to find a logically equivalent version where the negation is applied only to the BC predicate (not next to a quantifier).				
Making cha	nge example			
For which nonnega	tive integers n can we	make change for n with	a coins of value 5 cents	and 3 cents?
Restating: We can	make change for	, we cannot make	xe change for*	, and

Strong induction def

New! Proof by Strong Induction (Rosen 5.2 p337)

To prove that a universal quantification over the set of all integers greater than or equal to some base integer b holds, pick a fixed nonnegative integer j and then:

Basis Step: Show the statement holds for b, b + 1, ..., b + j.

Recursive Step: Consider an arbitrary integer n greater than or equal to b+j, assume (as the **strong**

induction hypothesis) that the property holds for each of $b, b + 1, \ldots, n$, and use

this and other facts to prove that the property holds for n+1.

Strong induction def

New! Proof by Strong Induction (Rosen 5.2 p337)

To prove that a universal quantification over the set of all integers greater than or equal to some base integer b holds, pick a fixed nonnegative integer j and then:

Basis Step: Show the statement holds for b, b + 1, ..., b + j.

Recursive Step: Consider an arbitrary integer n greater than or equal to b+j, assume (as the **strong**

induction hypothesis) that the property holds for each of $b, b + 1, \ldots, n$, and use

this and other facts to prove that the property holds for n+1.

Prime number def

Definition: An integer p greater than 1 is called **prime** if the only positive factors of p are 1 and p. A positive integer that is greater than 1 and is not prime is called composite.

Prime number def

Definition: An integer p greater than 1 is called **prime** if the only positive factors of p are 1 and p. A positive integer that is greater than 1 and is not prime is called composite.

Proof by contradiction def

New! Proof by Contradiction

To prove that a statement p is true, pick another statement r and once we show that $\neg p \to (r \land \neg r)$ then we can conclude that p is true.

Informally The statement we care about can't possible be false, so it must be true.

Rational number def

The set of rational numbers, \mathbb{Q} is defined as

$$\left\{\frac{p}{q}\mid p\in\mathbb{Z} \text{ and } q\in\mathbb{Z} \text{ and } q\neq 0\right\} \quad \text{ or, equivalently, } \quad \left\{x\in\mathbb{R}\mid \exists p\in\mathbb{Z} \exists q\in\mathbb{Z}^+ (p=x\cdot q)\right\}$$

Extra practice: Use the definition of set equality to prove that the definitions above give the same set.

Proof square root2

Goal: The square root of 2 is not a rational number. In other words: $\neg \exists x \in \mathbb{Q}(x^2 - 2 = 0)$

Attempted proof: The definition of the set of rational numbers is the collection of fractions p/q where p is an integer and q is a nonzero integer. Looking for a witness p and q, we can write the square root of 2 as the fraction $\sqrt{2}/1$, where 1 is a nonzero integer. Since the numerator is not in the domain, this witness is not allowed, and we have shown that the square root of 2 is not a fraction of integers (with nonzero denominator). Thus, the square root of 2 is not rational.

The problem in the above attempted proof is that

Proof:

Lemma 1: For every two integers p and q, not both zero, $gcd\left(\frac{p}{gcd(p,q)}, \frac{q}{gcd(p,q)}\right) = 1$.

Lemma 2: For every two integers a and b, not both zero, with gcd(a,b) = 1, it is not the case that both a is even and b is even.

Lemma 3: For every integer x, x is even if and only if x^2 is even.

Proposition def

Proposition Declarative sentence that is true or false (not both).

Propositional variable Variable that represents a proposition.

Compound proposition New propositions formed from existing propositions (potentially) using

logical operators.

Truth table Table with 1 row for each of the possible combinations of truth values

of the input and an additional column that shows the truth value of

the result of the operation corresponding to a particular row.

Note: A propositional variable is one example of a compound proposition.

Logical operators

Logical operators aka propositional connectives

Conjunction	AND	\wedge	$\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $	2 inputs	Evaluates to T when both inputs are T
Exclusive or	XOR	\oplus	\oplus	2 inputs	Evaluates to T when exactly one of inputs is T
Disjunction	OR	\vee	\lor	2 inputs	Evaluates to T when at least one of inputs is T
Negation	NOT	\neg	\label{lnot}	1 input	Evaluates to T when its input is F

Logical operators truth tables

Inp	out		Output			
		Conjunction	Exclusive or	Disjunction	Input	Output
p	q	$p \wedge q$	$p\oplus q$	$p \lor q$		Output Negation
\overline{T}	T	T	F	T	p	$\neg p$
T	F	F	T	T	T	\overline{F}
F	T	F	T	T	F	T
F	F	F	F	F	1	ı

Logical operators exercise

Input	Output				
p - q - r		$ \mid (p \wedge q) \oplus ((p \oplus q) \wedge r) $	$(p \wedge q) \vee ((p \oplus q) \wedge r)$		
T T T					
T T F					
T F T					
T F F					
F T T					
F T F					
F F T					
F F F					

Logical equivalence

Logical equivalence	Two compound propositions are logically equivalent means that they have the same truth values for all set-
Tautology	tings of truth values to their propositional variables. A compound proposition that evaluates to true for all settings of truth values to its propositional variables; it
Contradiction	is abbreviated T . A compound proposition that evaluates to false for all settings of truth values to its propositional variables; it
Contingency	is abbreviated F . A compound proposition that is neither a tautology nor a contradiction.

Logical equivalence exercise

Extra Example: Which of the compound propositions in the table below are logically equivalent?

Inp	out	Output					
p	q	$\neg (p \land \neg q)$	$\neg (\neg p \lor \neg q)$	$(\neg p \lor q)$	$(\neg q \lor \neg p)$	$(p \wedge q)$	
\overline{T}	T						
T	F						
F	T						
F	F						

Logical equivalence example

(Some) logical equivalences) cf. Rosen pp. 26-28

$$p \lor q \equiv q \lor p \qquad p \land q \equiv q \land p$$

$$(p \lor q) \lor r \equiv p \lor (q \lor r) \qquad (p \land q) \land r \equiv p \land (q \land r)$$

$$p \land F \equiv F \qquad p \lor T \equiv T \qquad p \land T \equiv p \qquad p \lor F \equiv p$$

$$\neg (p \land q) \equiv \neg p \lor \neg q \qquad \neg (p \lor q) \equiv \neg p \land \neg q$$

Commutativity Ordering of terms
Associativity Grouping of terms
Absorption aka short circuit evaluation
DeMorgan's Laws

Can replace p and q with any compound proposition

Truth table to compound proposition application

Application: design a circuit given a desired input-output relationship.

Inp	out	Out	put
p	q	$mystery_1$	$mystery_2$
\overline{T}	T	T	F
T	F	T	F
F	T	F	F
F	F	$\mid T$	T

1	npu	Output	
p	q	r	?
\overline{T}	T	T	T
T	T	F	T
T	F	T	F
T	F	F	T
F	T	T	F
F	T	F	F
F	F	T	T
F	F	F	$\mid F \mid$

A compound proposition that gives output $mystery_1$ is:	
A compound proposition that gives output $mystery_2$ is:	

Dnf cnf definition

Definition A compound proposition is in **disjunctive normal form** (DNF) means that it is an OR of ANDs of variables and their negations.

Definition A compound proposition is in **conjunctive normal form** (CNF) means that it is an AND of ORs of variables and their negations.

Extra example: A compound proposition that gives output? is:

Hypothesis conclusion

The only way to make the conditional statement $p \to q$ false is to				
The hypothesis of $p \to q$ is	The antecedent of $p \to q$ is			
The conclusion of $p \to q$ is	The consequent of $p \to q$ is			

Logical equivalence example two

Input	Output					
	Conjunction	Exclusive or	Disjunction	Conditional	Biconditional	
p - q	$p \wedge q$	$p\oplus q$	$p \lor q$	$p \rightarrow q$	$p \leftrightarrow q$	
T T	T	F	T	T	T	
T F	F	T	T	F	F	
F T	F	T	T	T	F	
F F	F	F	F	T	T	

Examples

Converse inverse contrapositive

The converse of $p \to q$ is	
The inverse of $p \to q$ is	 Which of these is logically equivalent to $p \to q$?
The contrapositive of $p \to q$ is	

Compound propositions translation

Translation: Express each of the following sentences as compound propositions, using the given propositions.

