

Black Box Testing

CS 240 – Advanced Programming Concepts

Black Box Testing

- Testing software against a specification of its external behavior without knowledge of internal implementation details
 - Can be applied to software “units” (e.g., classes) or to entire programs
 - External behavior is defined in API docs, Functional specs, Requirements specs, etc.
- Because black box testing purposely disregards the program's control structure, attention is focused primarily on the information domain (i.e., data that goes in, data that comes out)
- **The Goal:** Derive sets of input conditions (test cases) that fully exercise the external functionality

Black Box Testing

- Black box testing tends to find different kinds of errors than white box testing
 - Missing functions
 - Usability problems
 - Performance problems
 - Concurrency and timing errors
 - Initialization and termination errors
 - Etc.
- Unlike white box testing, black box testing tends to be applied later in the development process

The Information Domain: Inputs and Outputs

- Inputs
 - Individual input values
 - Try many different values for each individual input
 - Combinations of inputs
 - Individual inputs are not independent of each other
 - Programs process multiple input values together, not just one at a time
 - Try many different combinations of inputs in order to achieve good coverage of the input domain
 - Ordering and Timing of inputs
 - In addition to the particular combination of input values chosen, the ordering and timing of the inputs can also make a difference

The Information Domain: Inputs and Outputs

- Defining the input domain
 - Boolean value
 - T or F
 - Numeric value in a particular range
 - $99 \leq N \leq 99$
 - Integer, Floating point
 - One of a fixed set of enumerated values
 - {Jan, Feb, Mar, ...}
 - {Visa, MasterCard, Discover, ...}
 - Formatted strings
 - Phone numbers
 - File names
 - URLs
 - Credit card numbers
 - Regular expressions

Equivalence Partitioning

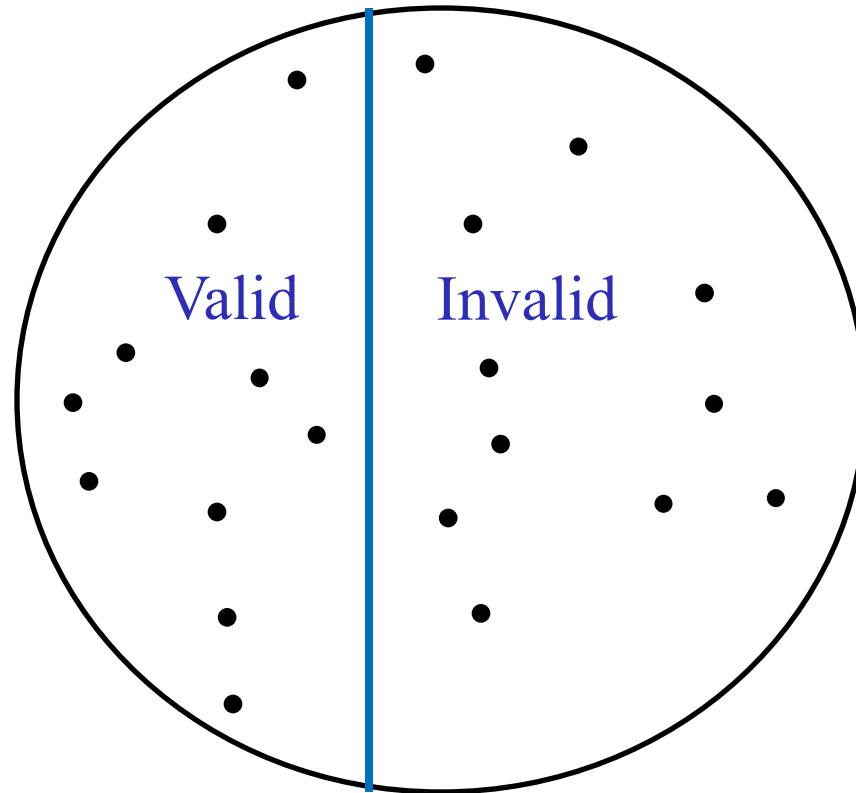
- Typically the universe of all possible test cases is so large that you cannot try them all
- You have to select a relatively small number of test cases to actually run
- Which test cases should you choose?
- Equivalence partitioning helps answer this question

Equivalence Partitioning

- Partition the test cases into "equivalence classes"
- Each equivalence class contains a set of "equivalent" test cases
- Two test cases are considered to be equivalent if we expect the program to process them both in the same way (i.e., follow the same path through the code)
- If you expect the program to process two test cases in the same way, only test a small number of them, thus reducing the number of test cases you have to run

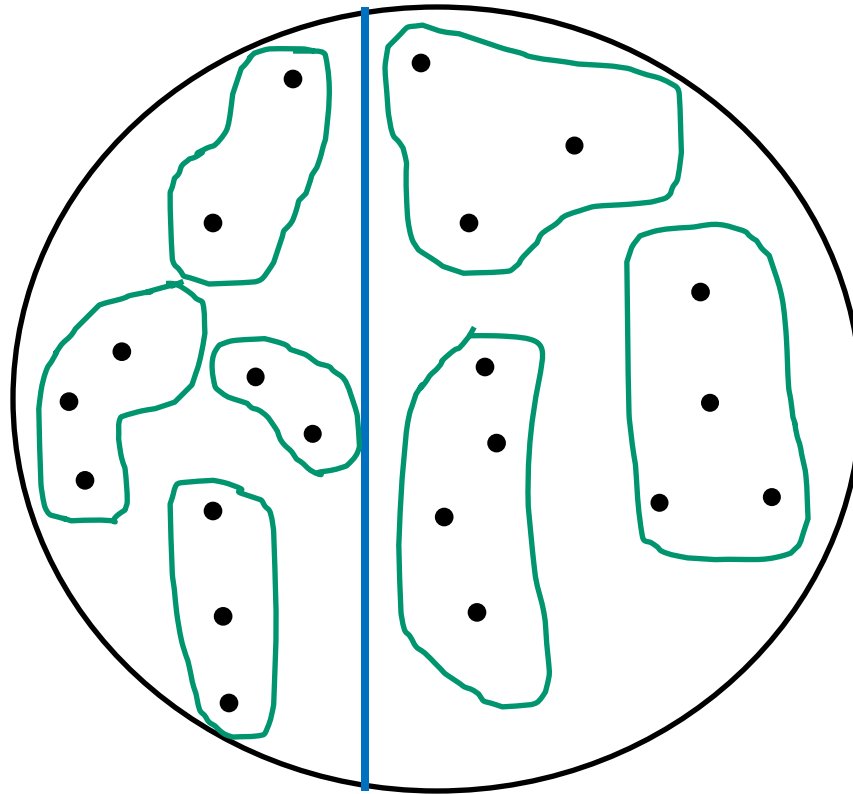
Equivalence Partitioning

- First-level partitioning: Valid vs. Invalid test cases



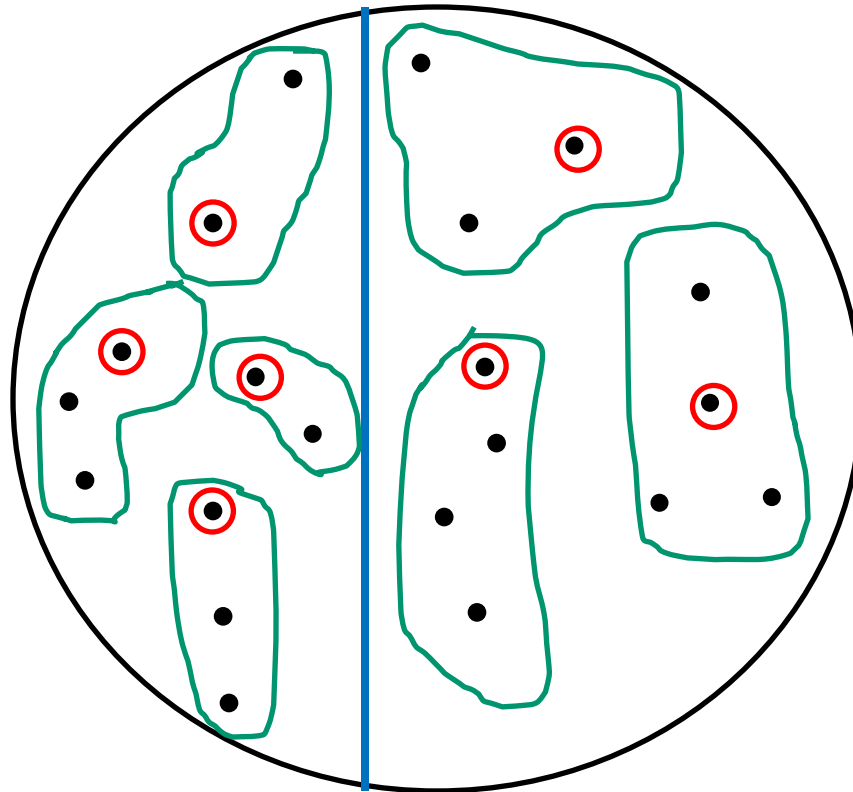
Equivalence Partitioning

- Partition valid and invalid test cases into equivalence classes



Equivalence Partitioning

- Create a test case for at least one value from each equivalence class



Equivalence Partitioning

- When designing test cases, you may use different definitions of “equivalence”, each of which will partition the test case space differently
- Example: `int add(int... value)`
 - Equivalence Definition 1: partition test cases by the number of inputs (1, 2, 3, etc.)
 - Equivalence Definition 2: partition test cases by the number signs they contain (positive, negative, both)
 - Equivalence Definition 3: partition test cases by the magnitude of operands (large numbers, small numbers, both)
 - Etc.

Equivalence Partitioning

- Example: `String fetch(URL)`
 - Equivalence Definition 1: partition test cases by URL protocol (“http”, “https”, “ftp”, “file”, etc.)
 - Equivalence Definition 2: partition test cases by type of file being retrieved (HTML, GIF, JPEG, Plain Text, etc.)
 - Equivalence Definition 3: partition test cases by length of URL (very short, short, medium, long, very long, etc.)
 - Etc.

Equivalence Partitioning

- **Oracle:** something that can tell you whether a test passed or failed
- If an oracle is available, the test values in each equivalence class can be randomly generated. This is more useful than always testing the same static values.
- Test multiple values in each equivalence class. Often you're not sure if you have defined the equivalence classes correctly or completely, and testing multiple values in each class is more thorough than relying on a single value.

Equivalence Partitioning - Examples

Input	Valid Equivalence Classes	Invalid Equivalence Classes
A integer N such that: -99 <= N <= 99	?	?
Phone Number Area code: [200, 999] Prefix: (200, 999] Suffix: Any 4 digits	?	?

[= inclusive

(= exclusive

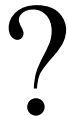
Equivalence Partitioning - examples

Input	Valid Equivalence Classes	Invalid Equivalence Classes
A integer N such that: $-99 \leq N \leq 99$	$[-99, -10]$ $[-9, -1]$ 0 $[1, 9]$ $[10, 99]$?
Phone Number Area code: $[200, 999]$ Prefix: $(200, 999]$ Suffix: Any 4 digits	?	?

Equivalence Partitioning - examples

Input	Valid Equivalence Classes	Invalid Equivalence Classes
A integer N such that: $-99 \leq N \leq 99$	$[-99, -10]$ $[-9, -1]$ 0 $[1, 9]$ $[10, 99]$	< -99 > 99 Malformed numbers $\{12-, 1-2-3, \dots\}$ Non-numeric strings $\{\text{junk}, 1E2, \$13\}$ Empty value
Phone Number Area code: $[200, 999]$ Prefix: $(200, 999]$ Suffix: Any 4 digits	?	?

Equivalence Partitioning - examples

Input	Valid Equivalence Classes	Invalid Equivalence Classes
A integer N such that: $-99 \leq N \leq 99$	$[-99, -10]$ $[-9, -1]$ 0 $[1, 9]$ $[10, 99]$	< -99 > 99 Malformed numbers $\{12-, 1-2-3, \dots\}$ Non-numeric strings $\{\text{junk}, 1E2, \$13\}$ Empty value
Phone Number Area code: $[200, 999]$ Prefix: $(200, 999]$ Suffix: Any 4 digits	555-5555 (555)555-5555 555-555-5555 $200 \leq \text{Area code} \leq 999$ $200 < \text{Prefix} \leq 999$	

Equivalence Partitioning - examples

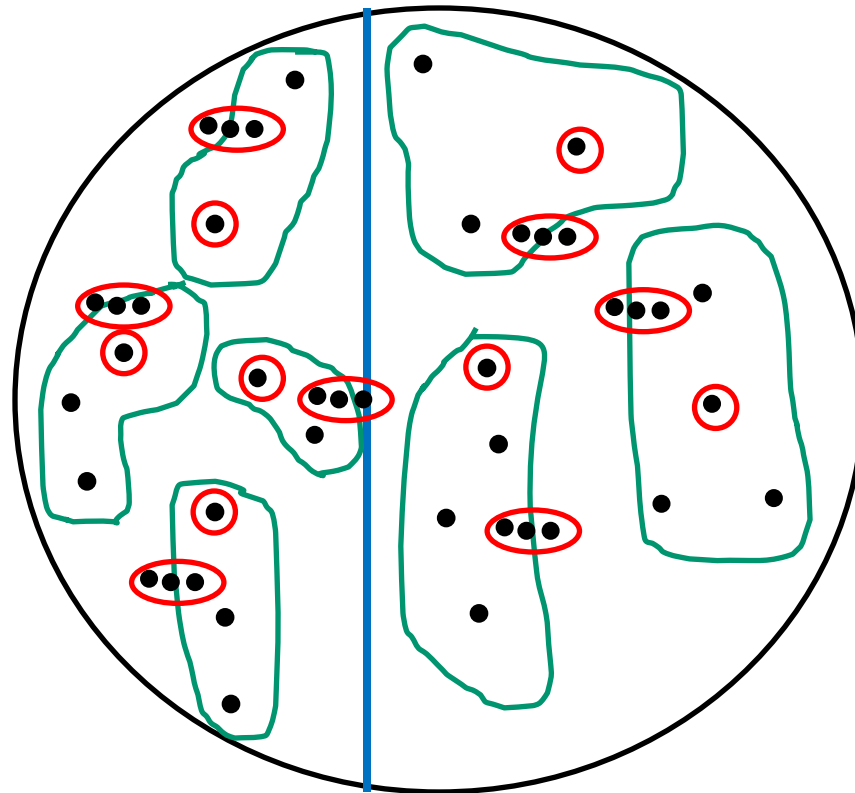
Input	Valid Equivalence Classes	Invalid Equivalence Classes
<p>A integer N such that: $-99 \leq N \leq 99$</p>	<p>$[-99, -10]$ $[-9, -1]$ 0 $[1, 9]$ $[10, 99]$</p>	<p>< -99 > 99 Malformed numbers $\{12-, 1-2-3, \dots\}$ Non-numeric strings $\{\text{junk}, 1E2, \\$13\}$ Empty value</p>
<p>Phone Number Area code: $[200, 999]$ Prefix: $(200, 999]$ Suffix: Any 4 digits</p>	<p>555-5555 (555)555-5555 555-555-5555 $200 \leq \text{Area code} \leq 999$ $200 < \text{Prefix} \leq 999$</p>	<p>Invalid format 5555555, (555)(555)5555, etc. Area code < 200 or > 999 Area code with non-numeric characters <i>Similar for Prefix and Suffix</i></p>

Boundary Value Analysis

- When choosing values from an equivalence class to test, use the values that are most likely to cause the program to fail
- Errors tend to occur at the boundaries of equivalence classes rather than at the "center"
 - `if (200 < areaCode && areaCode < 999) { // valid area code }`
 - Wrong!
 - `if (200 <= areaCode && areaCode <= 999) { // valid area code }`
 - Testing area codes 200 and 999 would catch this error, but a center value like 770 would not
- In addition to testing center values, we should also test boundary values
 - Right on a boundary
 - Very close to a boundary on either side

Boundary Value Analysis

- Create test cases to test boundaries of equivalence classes



Boundary Value Analysis - examples

Input	Boundary Cases
A number N such that: $-99 \leq N \leq 99$?
Phone Number Area code: [200, 999] Prefix: (200, 999] Suffix: Any 4 digits	?

Boundary Value Analysis - examples

Input	Boundary Cases
A number N such that: $-99 \leq N \leq 99$	-100, -99, -98 -10, -9 -1, 0, 1 9, 10 98, 99, 100
Phone Number Area code: [200, 999] Prefix: (200, 999] Suffix: Any 4 digits	?

Boundary Value Analysis - examples

Input	Boundary Cases
A number N such that: $-99 \leq N \leq 99$	-100, -99, -98 -10, -9 -1, 0, 1 9, 10 98, 99, 100
Phone Number Area code: [200, 999] Prefix: (200, 999] Suffix: Any 4 digits	Area code: 199, 200, 201 Area code: 998, 999, 1000 Prefix: 200, 199, 198 Prefix: 998, 999, 1000 Suffix: 3 digits, 5 digits

Boundary Value Analysis - examples

- Numeric values are often entered as strings which are then converted to numbers internally [int x = atoi(str);]
- This conversion requires the program to distinguish between digits and non-digits
- A boundary case to consider: Will the program accept / and : as digits?

Char	/	0	1	2	3	4	5	6	7	8	9	:
ASCII	47	48	49	50	51	52	53	54	55	56	57	58

Mainstream usage testing

- Don't get so wrapped up in testing boundary cases that you neglect to test "normal" input values
 - Values that users would typically enter during mainstream usage

Black Box Testing Examples

- Triangle Classification Example
- Next Date Example
- Substring Search Example

Ad Hoc Exploratory Testing (Error Guessing)

- Based on intuition, guess what kinds of inputs might cause the program to fail
- Create some test cases based on your guesses
- Intuition will often lead you toward boundary cases, but not always
- Some special cases aren't boundary values, but are mishandled by many programs
 - Try exiting the program while it's still starting up
 - Try loading a corrupted file
 - Try strange but legal URLs: `hTtP://Www.bYu.EDU/`

Comparison Testing (aka: A/B Testing)

- If you have multiple implementations of the same functionality, you can run test inputs through both implementations, and compare the results for equality
- Why would you have access to multiple implementations?
 - Safety-critical systems sometimes use multiple, independent implementations of critical modules to ensure the accuracy of results
 - You might use a competitor's product, or an earlier version of your own, as the second implementation
 - You might write a software simulation of a new chip that serves as the specification to the hardware designers. After building the chip, you could compare the results computed by the chip hardware with the results computed by the software simulator
- Inputs may be randomly generated or designed manually

Performance Testing

- Measure the system's performance
 - Running times of various tasks
 - Memory usage, including memory leaks
 - Network usage (Does it consume too much bandwidth? Does it open too many connections?)
 - Disk usage (Is the disk footprint reasonable? Does it clean up temporary files properly?)
 - Process/thread priorities (Does it play well with other applications, or does it hog the whole machine?)

Limit Testing

- Test the system at the limits of normal use
- Test every limit on the program's behavior defined in the requirements
 - Maximum number of concurrent users or connections
 - Maximum number of open files
 - Maximum request size
 - Maximum file size
 - Etc.
- What happens when you go slightly beyond the specified limits?
 - Does the system's performance degrade dramatically, or gracefully?

Stress Testing

- Test the system under extreme conditions (i.e., beyond the limits of normal use)
- Create test cases that demand resources in abnormal quantity, frequency, or volume
 - Low memory conditions
 - Disk faults (read/write failures, full disk, file corruption, etc.)
 - Network faults
 - Unusually high number of requests
 - Unusually large requests or files
 - Unusually high data rates (what happens if the network suddenly becomes ten times faster?)
- Even if the system doesn't need to work in such extreme conditions, stress testing is an excellent way to find bugs

Testing for Race Conditions and Other Timing Dependencies

- Many systems perform multiple concurrent activities
 - Operating systems manage concurrent programs, interrupts, etc.
 - Servers service many clients simultaneously
 - Applications let users perform multiple concurrent actions
- Test a variety of different concurrency scenarios, focusing on activities that are likely to share resources (and therefore conflict)
- "Race conditions" are bugs that occur only when concurrent activities interleave in particular ways, thus making them difficult to reproduce
- Test on hardware of various speeds to ensure that your system works well on both slower and faster machines

Random Testing

- Randomly generate test inputs
 - Could be based on some statistical model
- How do you tell if the test case succeeded?
 - Where do the expected results come from?
 - Some type of “oracle” is needed
- Expected results could be calculated manually
 - Possible, but lots of work
- Automated oracles can often be created to measure characteristics of the system
 - Performance (memory usage, bandwidth, running times, etc.)
 - Did the system crash?
 - Maximum and average user response time under simulated user load

Security Testing

- Any system that manages sensitive information or performs sensitive functions may become a target for intrusion (i.e., hackers)
- How feasible is it to break into the system?
- Learn the techniques used by hackers
- Try whatever attacks you can think of
- Hire a security expert to break into the system
- If somebody broke in, what damage could they do?
- If an authorized user became disgruntled, what damage could they do?

Usability Testing

- Is the user interface intuitive, easy to use, organized, logical?
- Does it frustrate users?
- Are common tasks simple to do?
- Does it conform to platform-specific conventions?

- Get real users to sit down and use the software to perform some tasks
 - Watch them performing the tasks, noting things that seem to give them trouble
 - Get their feedback on the user interface and any suggested improvements

- Report bugs for any problems encountered

Recovery Testing

- Try turning the power off or otherwise crashing the program at arbitrary points during its execution
 - Does the program come back up correctly when you restart it?
 - Was the program's persistent data corrupted (files, databases, etc.)?
 - Was the extent of user data loss within acceptable limits?
- Can the program recover if its configuration files have been corrupted or deleted?
- What about hardware failures? Does the system need to keep working when its hardware fails? If so, verify that it does so.

Configuration Testing

- Test on all required hardware configurations
 - CPU, memory, disk, graphics card, network card, etc.
- Test on all required operating systems and versions thereof
 - Virtualization technologies such as VMWare and Virtual PC are very helpful for this
- Test as many Hardware/OS combinations as you can
- Test installation programs and procedures on all relevant configurations

Compatibility Testing

- Test to make sure the program is compatible with other programs it is supposed to work with
- Ex: Can Word 12.0 load files created with Word 11.0?
- Ex: "Save As... Word, Word Perfect, PDF, HTML, Plain Text"
- Ex: "This program is compatible with Internet Explorer and Firefox"
- Test all compatibility requirements

Documentation Testing

- Test all instructions given in the documentation to ensure their completeness and accuracy
- For example, “How To ...” instructions are sometimes not updated to reflect changes in the user interface
- Test user documentation on real users to ensure it is clear and complete