Applied Machine Learning

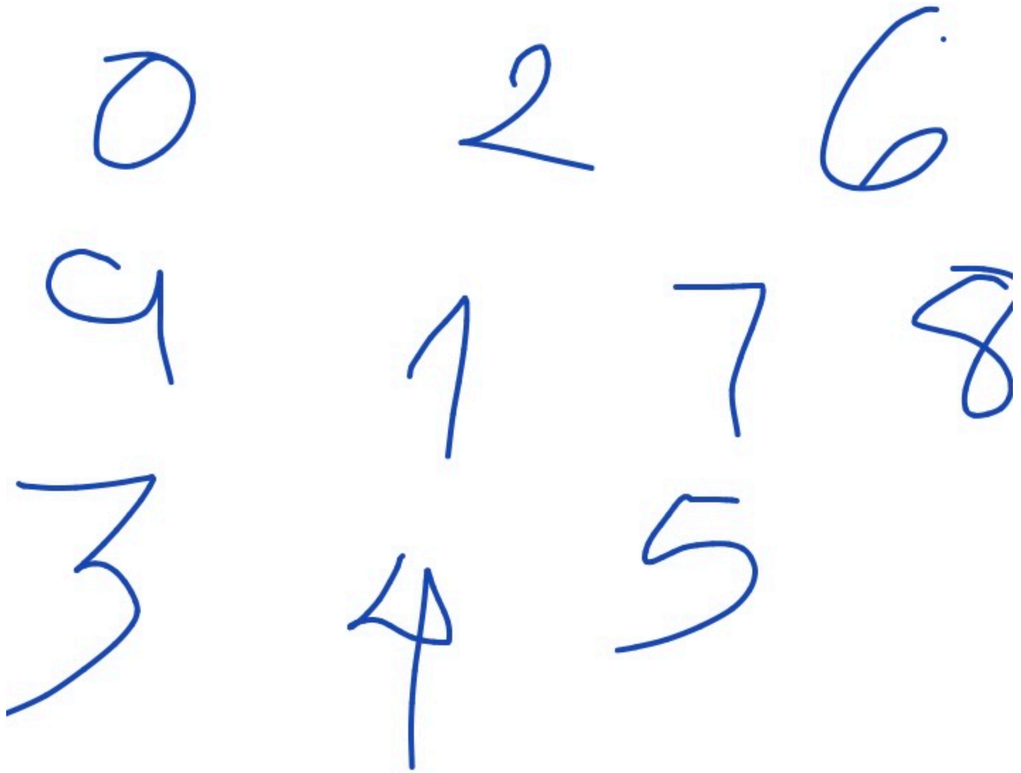
Classification - Introduction

Classification - Introduction

- Classifiers
- Binary classification and multi-class classification
- How to measure how good is a classifier

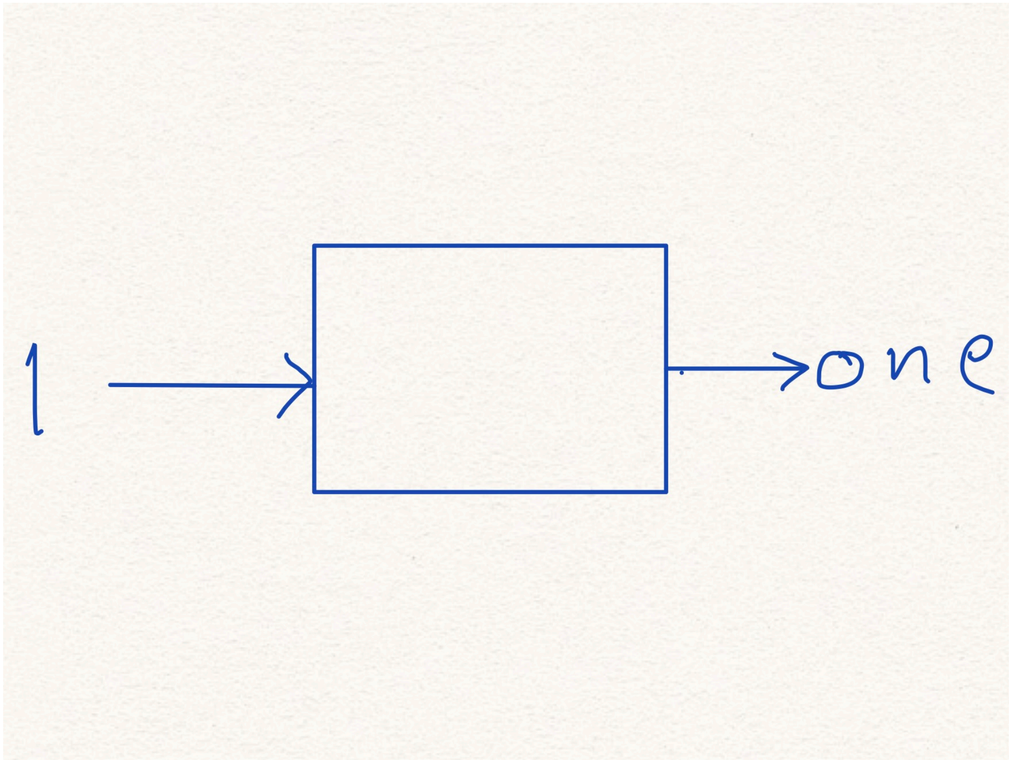
Teaching to read digits

- Show the student a sample of written digits, one at a time
  - Training Set
- Shuffle the images and evaluate the student
- Repeat until the student gets most of them right
- The student tries with a new set
  - Test Set



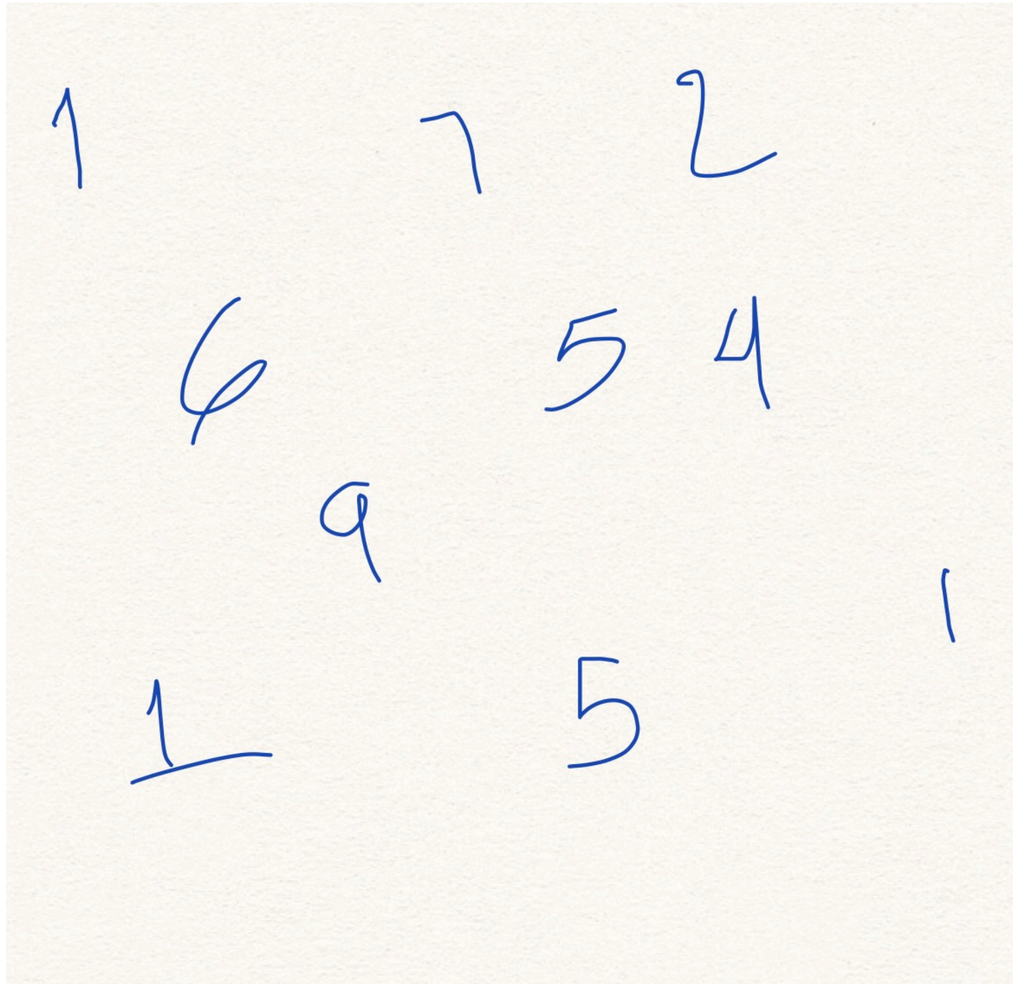
Binary Classification

- Assign each example to a class
  - it is the digit one
  - it is not the digit one



Evaluating the classifier

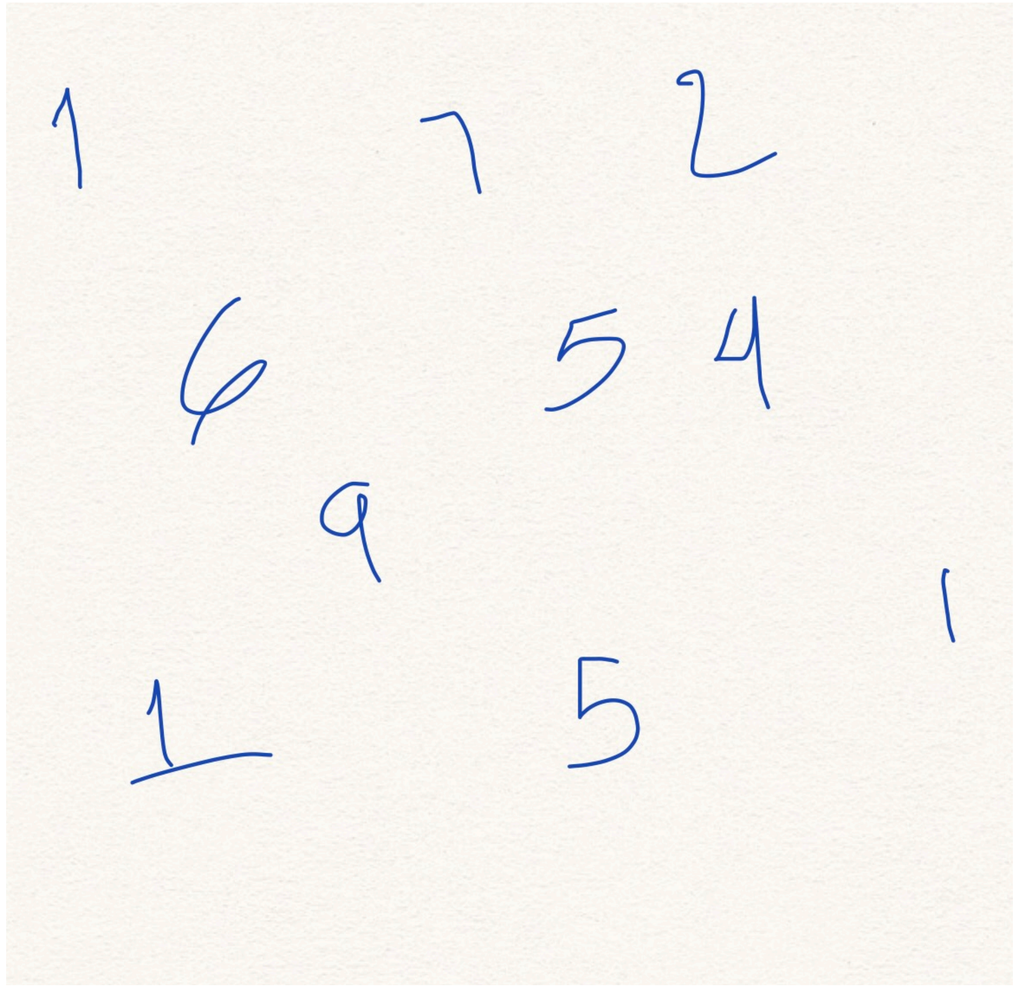
- Correctly identifies a cases of 1 as ONE
  - True Positive
- Incorrectly identifies a case of 7 as ONE
  - False Positive
- Incorrectly identifies a case of 1 as NOT A ONE
  - False Negative
- Correctly identifies a case of 5 as NOT A ONE
  - True Negative



Evaluating the classifier

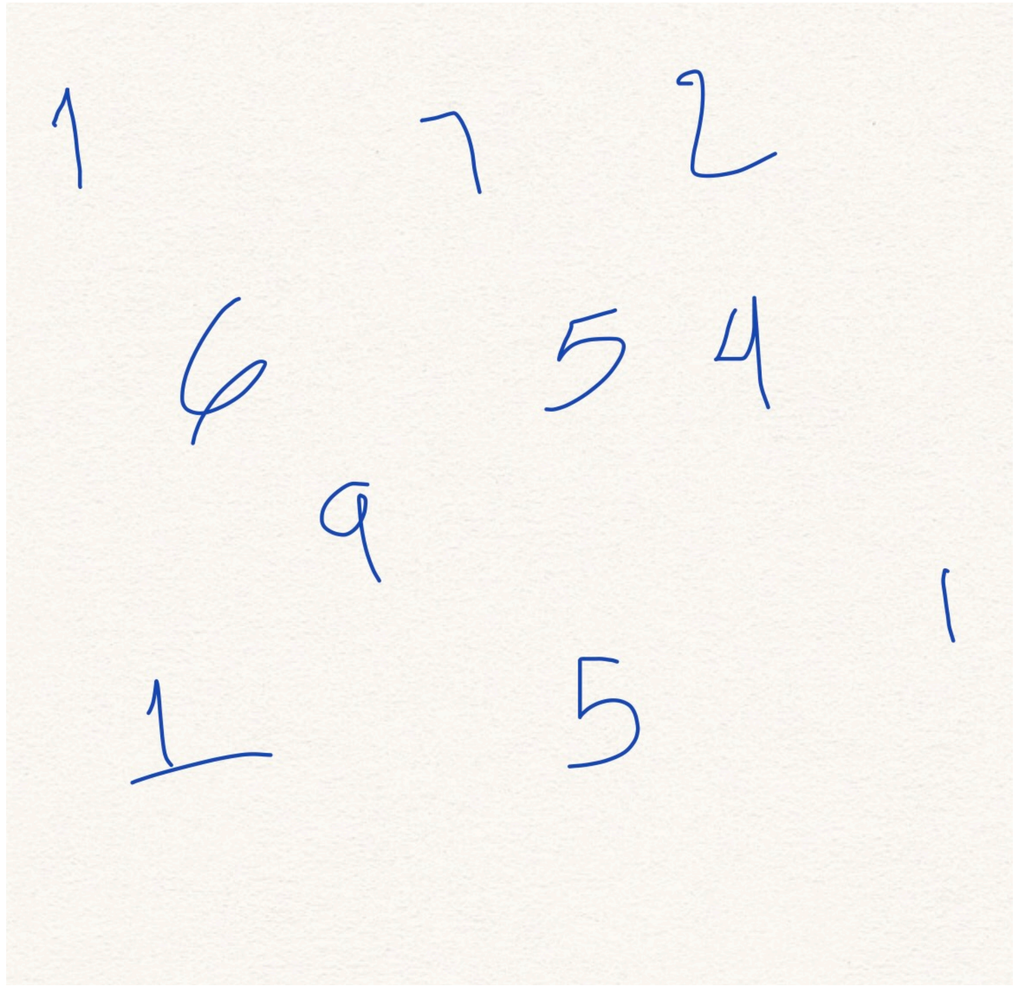
True Label	Student identifies	
	ONE	NOT A ONE
ONE	1, 1	1
NOT A ONE	7	6, 5, 5, 2, 4, 9

- **Overall Accuracy: 8/10**
- **Error Rate: 2/10**



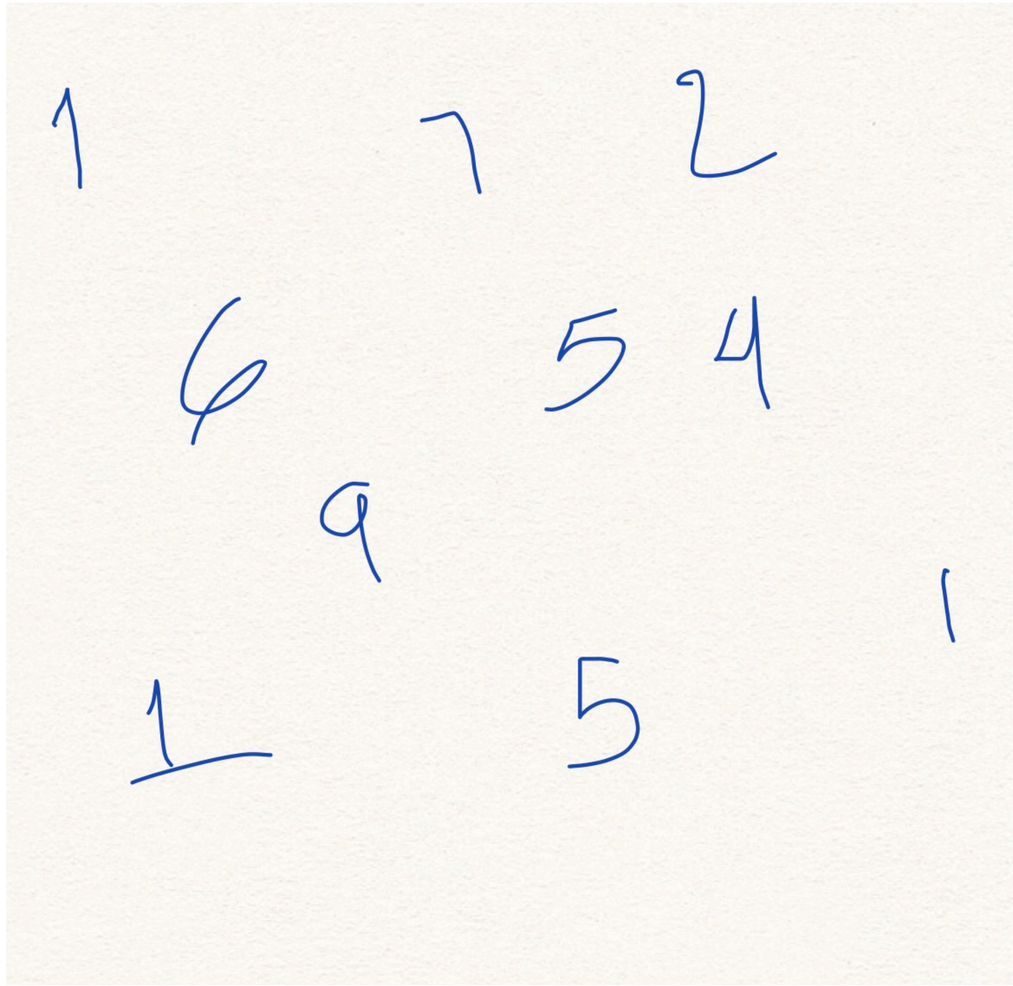
Evaluating the classifier

True Label	Student identifies		Correct	Incorrect
	ONE	NOT A ONE	Accuracy	Error rate
ONE	1, 1	1	2/3	1/3
NOT A ONE	7	6, 5, 5, 2, 4, 9	6/7	1/7



Evaluating the classifier

True Label	Student identifies		Correct	Incorrect
	ONE	NOT A ONE	Accuracy	Error rate
ONE	1, 1	1	2/3	1/3
NOT A ONE	7	6, 5, 5, 2, 4, 9	6/7	1/7

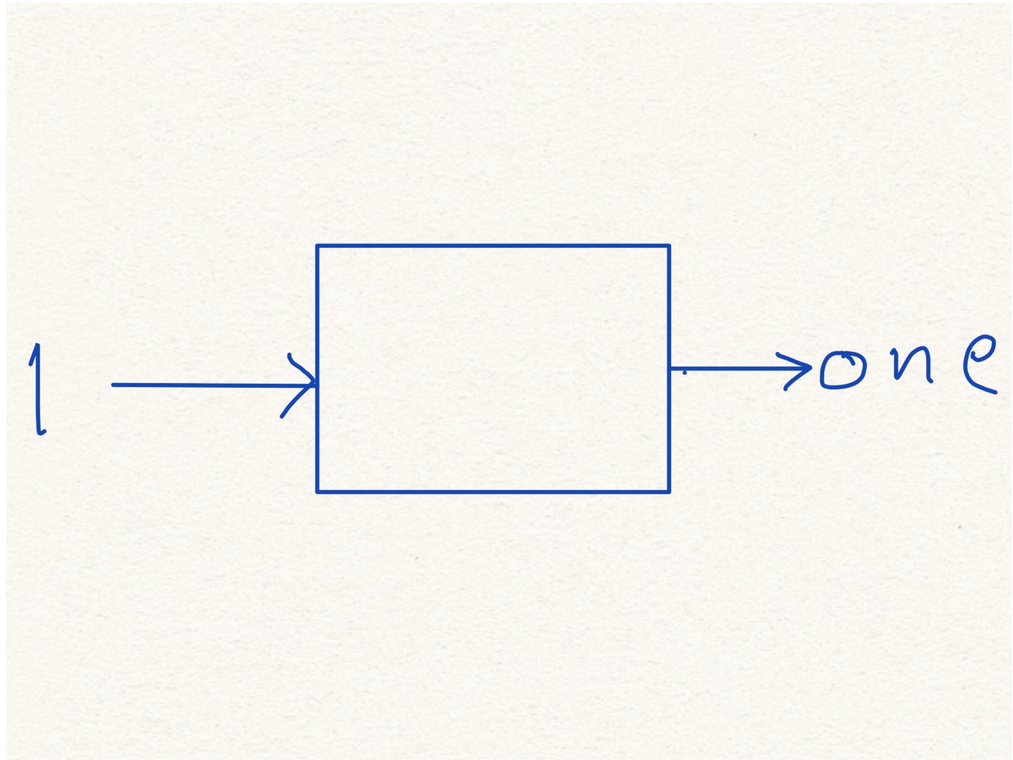


- Sensitivity $= \frac{TP}{TP + FN} = \frac{2}{3}$

- Specificity $= \frac{TN}{TN + FP} = \frac{6}{7}$

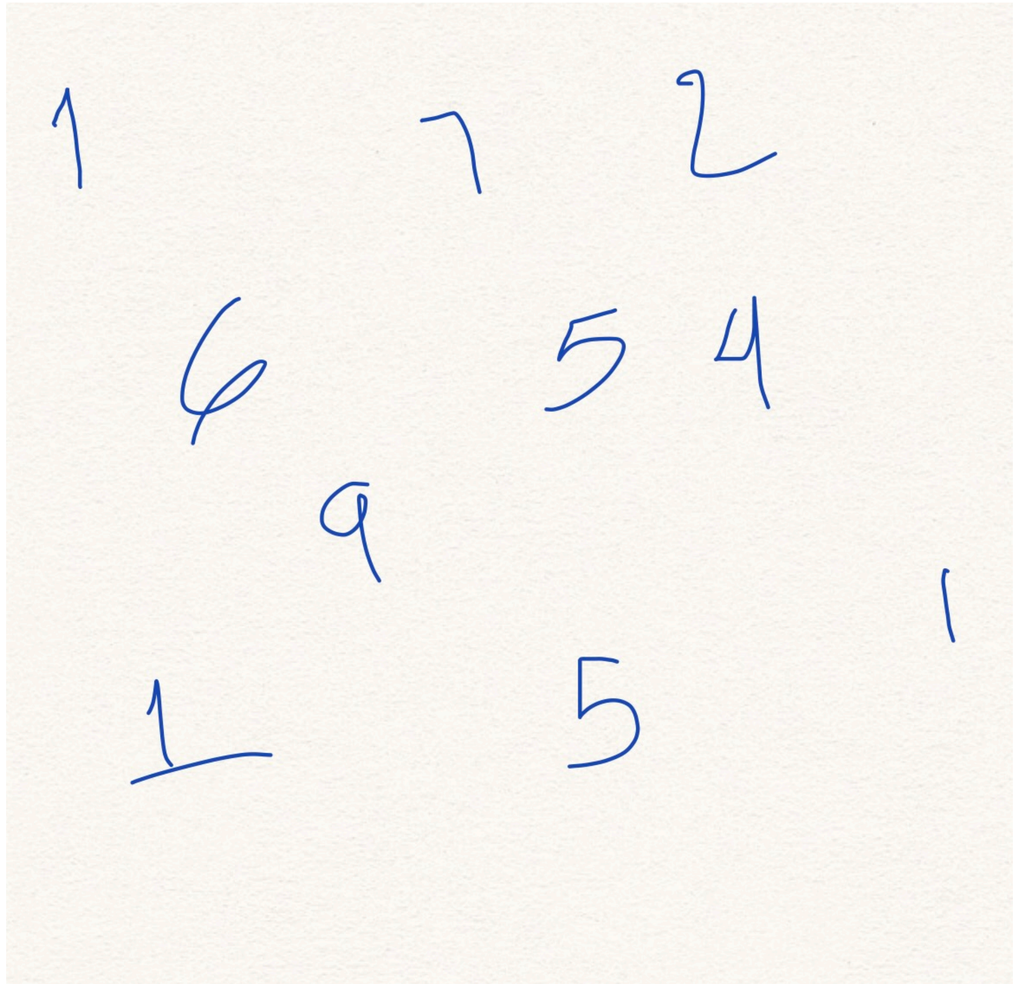
Multi-Class Classification

- Assign each example to a class
  - ZERO, ONE, TWO, …, NINE



Class-Confusion Matrix

True Label	Student identifies										Class Error %
	"0"	"1"	"2"	"3"	"4"	"5"	"6"	"7"	"8"	"9"	
"0"	9									1	10
"1"		7						3			30
"2"		1	8					1			20
"3"				9					1		10
"4"					7					3	30
"5"						10					0
"6"	1						9				10
"7"			1					9			10
"8"									10		0
"9"									1	6	40



- **Number of examples in the set = 100**
- **Accuracy = 84 %**

Training and Testing

- Collect all the available examples and split them into
  - Training Set
  - Test Set for Evaluation
- Potential issue:
  - Overfitting or Selection Bias

Cross-Validation

- Iteratively
  - generate a new random split, or Fold, of the Training Set and Test Set
  - teach with the Training Set
  - evaluate with the Test Set and record accuracy for current Fold
- Average accuracy over all the Folds

Leave-One-Out Cross-Validation

- A particular case of cross-validation with Sizes as follows:
  - Data set: N
  - Training Set in each Fold: N-1
  - Test Set in each Fold: 1

k-Fold Cross-Validation

- A particular case of cross-validation with Sizes as follows:
  - Data set: N
    - Split into k equal pieces at fixed places
    - | N/K | N/K | … | N/K |
  - Training Set in each Fold: N-N/K
  - Test Set in each Fold: N/K
    - | Test | N/K | … | N/K |
    - | N/K | Test | … | N/K |

Classifiers

- Input: data item
  - set of features in vector X
- Output: label
  - Y
    - Binary
    - Multi-class

Classifiers

- Supervised training
  - Input: Training Set: data items pairs (X, Y)
- Testing
  - Input: Test Set: data items (X)
  - Evaluate performance on Y

Classifiers

- Goal: perform well on Test Set
  - Class-Confussion Matrix
  - Accuracy
  - Error rate
  - Sensitivity
  - Sensibility

Applications

- Medicine
- Wheather
- Entertainment
- …

Classification - Introduction

- Classification
- Types of classifiers
- Training and Testing
- Evaluation