Machine Learning – I

Richa Singh

Google classroom code: wgzuohn

Slides are prepared from several information sources on the web and books

About the instructors

- Richa Singh:
  - Professor, IIT Jodhpur
  - Faculty, IIIT Delhi (2009 – 2019)
  - home.iitj.ac.in/~richa

  - Adjunct Professor, West Virginia University
  - Vice President – Publications, IEEE Biometrics Council
  - Associate Editor-in-Chief, Pattern Recognition Journal

About the instructors

- Dr. Rajendra Nagar
  - Assistant Professor, EE
  - home.iitj.ac.in/~rn

- Dr. Anand Mishra
  - Assistant Professor, CSE
  - home.iitj.ac.in/~mishra

Teaching Assistants

- Abhiram
- Harsha M.
- Kartik Thakral
- Neelu Verma
- Surbhi Mittal
- Vikansh Nath

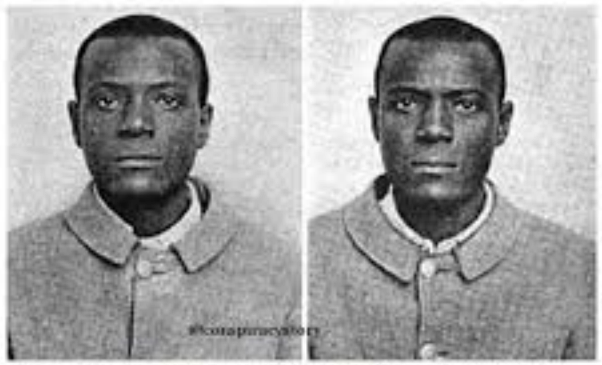
MTech and PhD students

Applications of ML

What are the facial expressions?

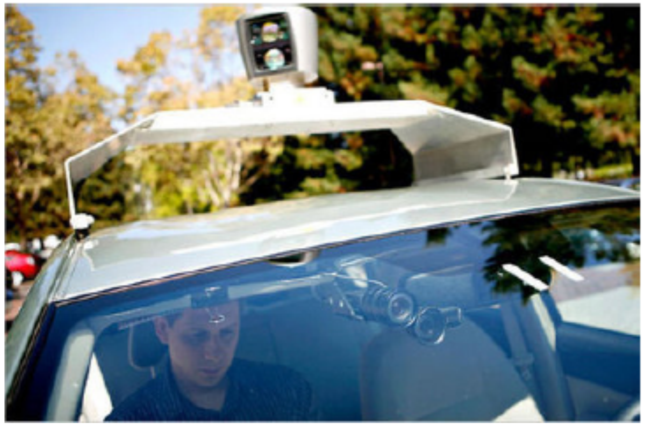
Weather forecast

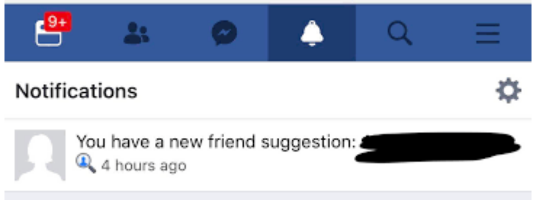
What are these letters?



Are they same or different?

Applications of ML





Background

- What is your background?

Please fill the questionnaire shared on WebEx chat.

Course Objectives

- The students are expected to have the ability to:
  - Develop a sense of Machine Learning in the modern context, and independently work on problems relating to Machine Learning
  - Design and program efficient algorithms related to Machine Learning, train models, conduct experiments, and deliver ML-based applications

Course Content

- Introduction: Motivation, Different types of learning, Linear regression, Logistic regression
- Decision Trees: Decision Tree algorithms, Random forests
- Model selection and validation: Validation for model selection, k-fold cross-validation, Training-Validation-Testing split, Regularized loss minimization
- Support Vector Machines: Hard SVM, Soft SVM, Optimality conditions, Duality, Kernel trick, Implementing Soft SVM with Kernels
- Neural Networks: Feedforward neural networks, Expressive power of neural networks, SGD and Backpropagation
- Gradient Descent: Introduction, Stochastic Gradient Descent, Subgradients, Stochastic, Gradient Descent for risk minimization

Course Content

- Nearest Neighbor: k-nearest neighbor, Curse of dimensionality
- Clustering: Linkage-based clustering algorithms, k-means algorithm, Spectral clustering
- Dimensionality reduction: Principal Component Analysis, Random projections, Compressed sensing
- Generative Models: Maximum likelihood estimator, Naive Bayes, Linear Discriminant Analysis, Latent variables and Expectation-maximization algorithm, Bayesian learning
- Feature Selection and Generation: Feature selection, Feature transformations, Feature learning
- Statistical Learning Framework: PAC learning, Agnostic PAC learning, Bias-complexity tradeoff, No free lunch theorem, VC dimension, Structural risk minimization, Boosting
- Foundations of Deep Learning: DNN, CNN, RNN, Autoencoders

Pre-requisites

- Programming
- Probability and Statistics
- Linear Algebra (matrices and vectors)

- Revise: resources will be uploaded on classroom

- Assignment zero (will be announced this week)
- Will be evaluated − bonus assignment

Class Resources

- Google classroom: wgzuohn
  - Announcements
  - Assignment posting and submission
  - Discussions
- Mailing list: mtech_exec_ML@iitj.ac.in

- Communications channel:
  - Email
  - Google classroom

Reading Resources

- Textbook:
  - Shalev-Shwartz,S., Ben-David,S., (2014), Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press
  - Tom Mitchell, Machine Learning
- Reference Books:
  - C. Bishop, Pattern Recognition and Machine Learning, Springer
  - K. Murphy, Machine Learning: a Probabilistic Perspective, MIT Press

Evaluation Components

- Grading
  - Assignments (programming and written): 30%
  - Exams: 40%
  - Project: 20%
  - Quiz : 10%
- Project team size: 2 students
  - Predefined project topics: you have to select one
- Assignments: individually

Collaboration Policy

- Discussion with friends and colleagues is good… but
  - the objective should be to improve understanding and learning
  - Not getting answers
- If you have discussed with anyone, you should acknowledge who helped you − from the class or outside the class
- Academic dishonesty policy of IITJ will apply

Any questions regarding administrative guidelines?

Please ask via Webex chat

Machine Learning

- What do we understand by learning?
  – Learning is any process by which a system improves performance from experience."

–Herbert Simon (1950)

- Machine Learning is the study of algorithms that
  – improve their performance P
  – at some task T
  – with experience E.

  A well-defined learning task is given by <P, T, E>.

- Tom Mitchell (1998)

Task, T

- Classification
- Regression
- Ranking
- Recommendation
- Density estimation
- ….

Performance, P

- Metric used to evaluate the performance of T
  - Classification: error rate or accuracy
  - Regression: mean squared error
  - Density estimation: probability assigned to samples

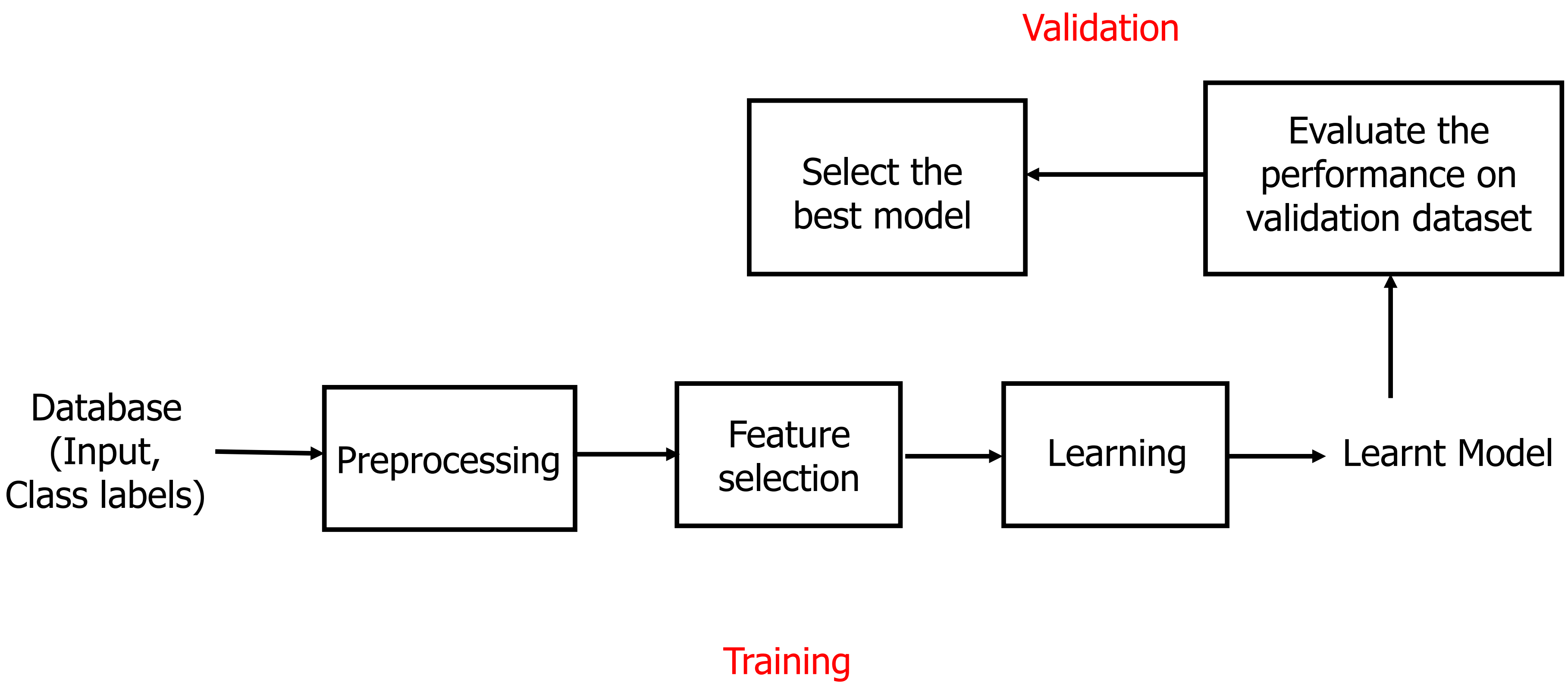
Machine Learning Pipeline

- Three steps:
  - Training
  - Validation
  - Testing

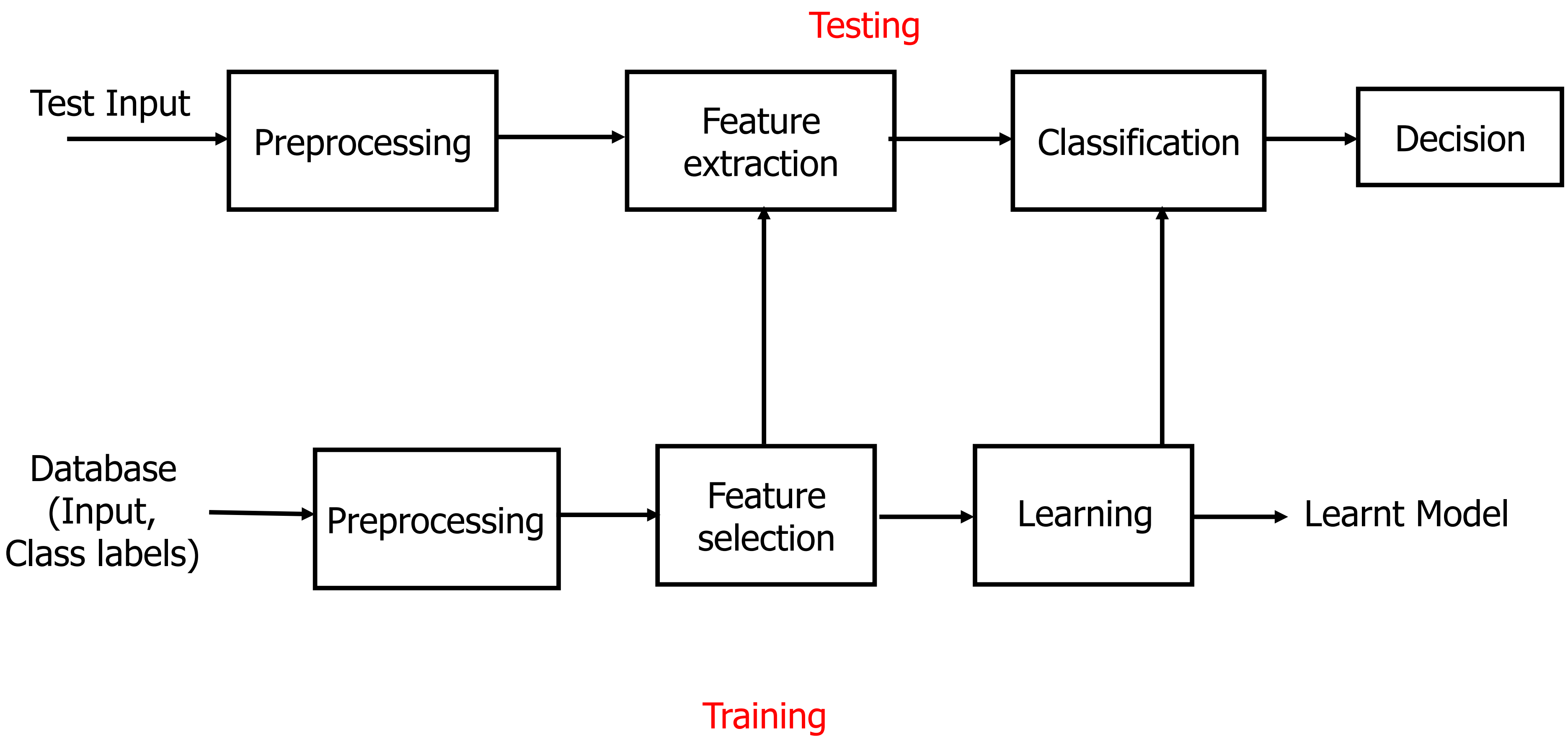
Machine Learning Pipeline

Training

Machine Learning Pipeline



Machine Learning Pipeline



ML in Practice

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

Based on a slide by Pedro Domingos

Types of ML Paradigms

- Supervised learning
  - Classification
  - Regression
- Unsupervised learning
- Reinforcement learning

Supervised Learning

- Given: training data + desired outputs (labels)
- (x1, y1), (x2, y2), ..., (xn, yn)
- Learn a function f(x) to predict y given x

Cats

Dogs

Supervised Learning

Child

Adult

Classification

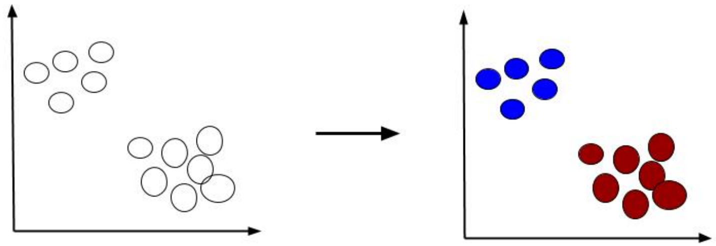
6 years

?? years

Regression

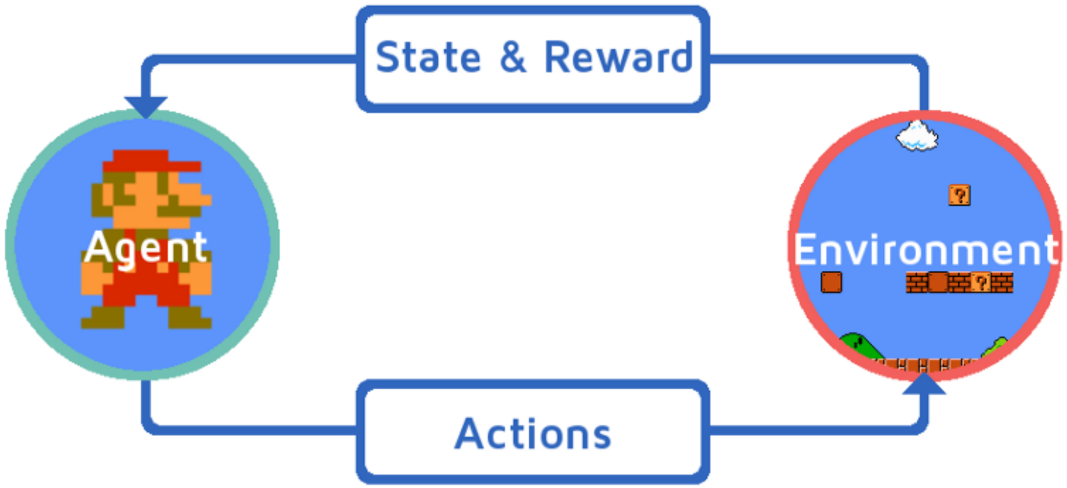
Unsupervised Learning

- Given: training data (without desired outputs)
- x1, x2, ..., xn (without labels)
- Output hidden structure behind the x's

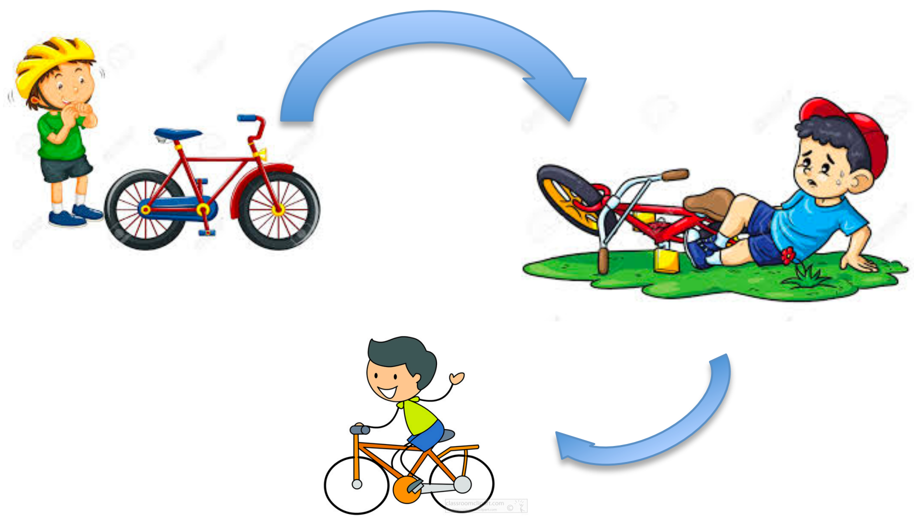


Clustering

Reinforcement Learning



Rewards from sequence of actions



Applications of Learning Paradigms

Supervised
- Person identification
- Object recognition
- Stock prediction

Reinforcement
- Game playing
- Credit assignment

Unsupervised
- Social network analysis
- Dimensionality reduction
- Market segmentation

Some ML Algorithms

- Supervised learning
  - Decision tree induction
  - Linear regression
  - Logistic regression
  - Support vector machines & kernel methods
  - Model ensembles
  - Bayesian learning
  - Neural networks & deep learning
  - Learning theory
- Unsupervised learning
  - Clustering
  - Dimensionality reduction
- Reinforcement learning
  - Temporal difference
  - Learning
  - Q learning

History of ML

- 1950s
  - Samuel's checker player
  - Selfridge's Pandemonium
- 1960s:
  - Neural networks: Perceptron
  - Pattern recognition
  - Learning in the limit theory
  - Minsky and Papert prove limitations of Perceptron
- 1970s:
  - Symbolic concept induction
  - Winston's arch learner
  - Expert systems and the knowledge acquisition bottleneck
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis
  - Scientific discovery with BACON
  - Mathematical discovery with AM

History of ML

- 2000s
  - Support vector machines & kernel methods
  - Graphical models
  - Statistical relational learning
  - Transfer learning
  - Sequence labeling
  - Collective classification and structured outputs
  - Computer Systems Applications (Compilers, Debugging, Graphics, Security)
  - E-mail management
  - Personalized assistants that learn
  - Learning in robotics and vision
- 2010s
  - Deep learning systems
  - Learning for big data
  - Bayesian methods
  - Multi-task & lifelong learning
  - Applications to vision, speech, social networks, learning to read, etc.
- 2020s
  - Deep learning
  - ???

Based on slide by Ray Mooney

Any questions? Please ask on Google Classroom

Questionnaire shared with you

Let us design a simple classification algorithm

Purse vs Laptop Bag: Design a classifier








Laptop bag vs. Purse: Design a classifier

- Features:
  - Width
  - Height
- Classifier: threshold

Evaluation Metrics

Let the problem statement be: classifying purse and bags.
Purses are labeled as positive class and bags are labeled as negative class

		Predicted Class	
		Negative	Positive
Actual Class	Negative	A (true negative)	C (false positive)
	Positive	D (false negative)	B (true positive)

Term	Meaning	Example
True positive	Correct classification	Purse identified as purse
False positive	Incorrect classification	Bag identified as purse
True negative	Correct classification	Bag identified as bag
False negative	Incorrect classification	Purse identified as bag

http://www.cs.rpi.edu/~leen/misc-publications/SomeStatDefs.html

Evaluation Metrics

		Predicted Class	
		Negative	**Positive**
Actual Class	**Negative**	A (true negative)	C (false positive)
	Positive	D (false negative)	B (true positive)

Metric	**Formula**
Average classification accuracy	TN / (TN + FP) + TP / (TP + FN)
Type I error (false positive rate)	FP / (TN + FP)
Type II error (false negative rate)	FN / (FN + TP)
True positive rate	TP / (TP + FN)
True negative rate	TN / (TN + FP)

http://www.cs.rpi.edu/~leen/misc-publications/SomeStatDefs.html

Evaluation Metrics

Metric	Formula
Average classification accuracy	(TN + TP) / (TN+TP+FN+FP)
Type I error (false positive rate)	FP / (TN + FP)
Type II error (false negative rate)	FN / (FN + TP)
True positive rate	TP / (TP + FN)
True negative rate	TN / (TN + FP)

- Type I error or false positive rate: The chance of incorrectly classifying a (randomly selected) sample as positive
- Type II error or false negative rate: The chance of incorrectly classification a (randomly selected) sample as negative

Evaluation Metrics

Metric	Formula
Average classification accuracy	(TN + TP) / (TN+TP+FN+FP)
Type I error (false positive rate)	FP / (TN + FP)
Type II error (false negative rate)	FN / (FN + TP)
True positive rate	TP / (TP + FN)
True negative rate	TN / (TN + FP)

- Type I error or false positive rate: The chance of incorrectly classifying a (randomly selected) sample as p
- Type II error or false negative rate: The cha classification a (randomly selected) samp

Prevalent in computer vision and image processing related classification problems

Any questions? Solve the question shared via chat.

Please ask on WebEx chat

Evaluation Metrics

Metric	Formula
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)

Precision: Fraction of retrieved instances that are relevant

Recall: Fraction of relevant instances that are retrieved

Evaluation Metrics

Metric	Formula
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)

Precision: Probability that a (randomly selected) retrieved document in relevant

Recall: Probability that a (randomly selected) relevant document is retrieved in a search

Evaluation Metrics

Metric	Formula
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)

Precision: Probability that a (randomly selected) r
document in relevant

Recall: Probability that a (randomly selected) re
document is retrieved in a search

More prevalent in information retrieval domain

Evaluation Metrics

Metric	Formula
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
Predictive value for a positive result (PV+)	TP / (TP + FP)
Predictive value for a negative result (PV-)	TN / (TN + FN)

Sensitivity: Proportion of actual positives which are correctly identified

Specificity: Proportion of actual negatives which are correctly identified

Evaluation Metrics

Metric	Formula
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
Predictive value for a positive result (PV+)	TP / (TP + FP)
Predictive value for a negative result (PV-)	TN / (TN + FN)

Sensitivity: The chance of correctly identifying positive samples. A sensitive test helps rule out disease (when the result is negative)

Specificity: The chance of correctly classifying negative samples. A very specific test rules in disease with a higher degree of confidence.

More prevalent in research related to medical sciences

Evaluation Metrics

Metric	Formula
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
Predictive value for a positive result (PV+)	TP / (TP + FP)
Predictive value for a negative result (PV-)	TN / (TN + FN)

Predictive value of a positive result: If the test is positive, what is the probability that the patient actually has the disease

Predictive value of a negative result: If the test is negative, what is the probability that the patient does not have the disease

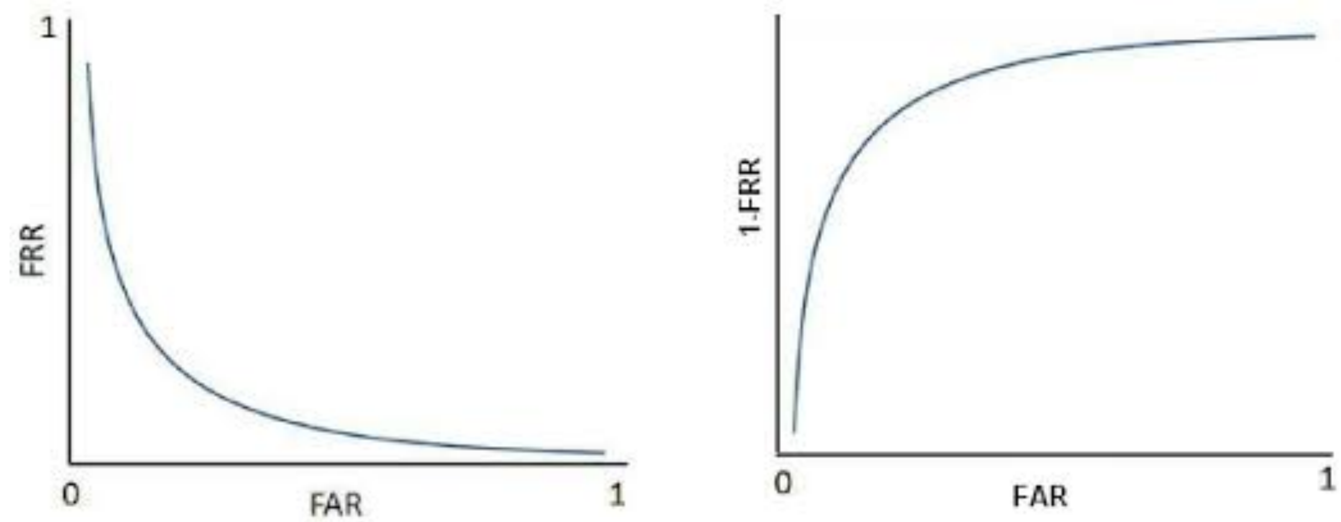
Performance Evaluation

- Classification is of two types:
  - Authentication / verification (1:1 Matching)
    - Is she Richa?
    - Is this an image of a helicopter?
  - Identification (1:n matching)
    - Who’s photo is this?
    - This image belongs to which class?

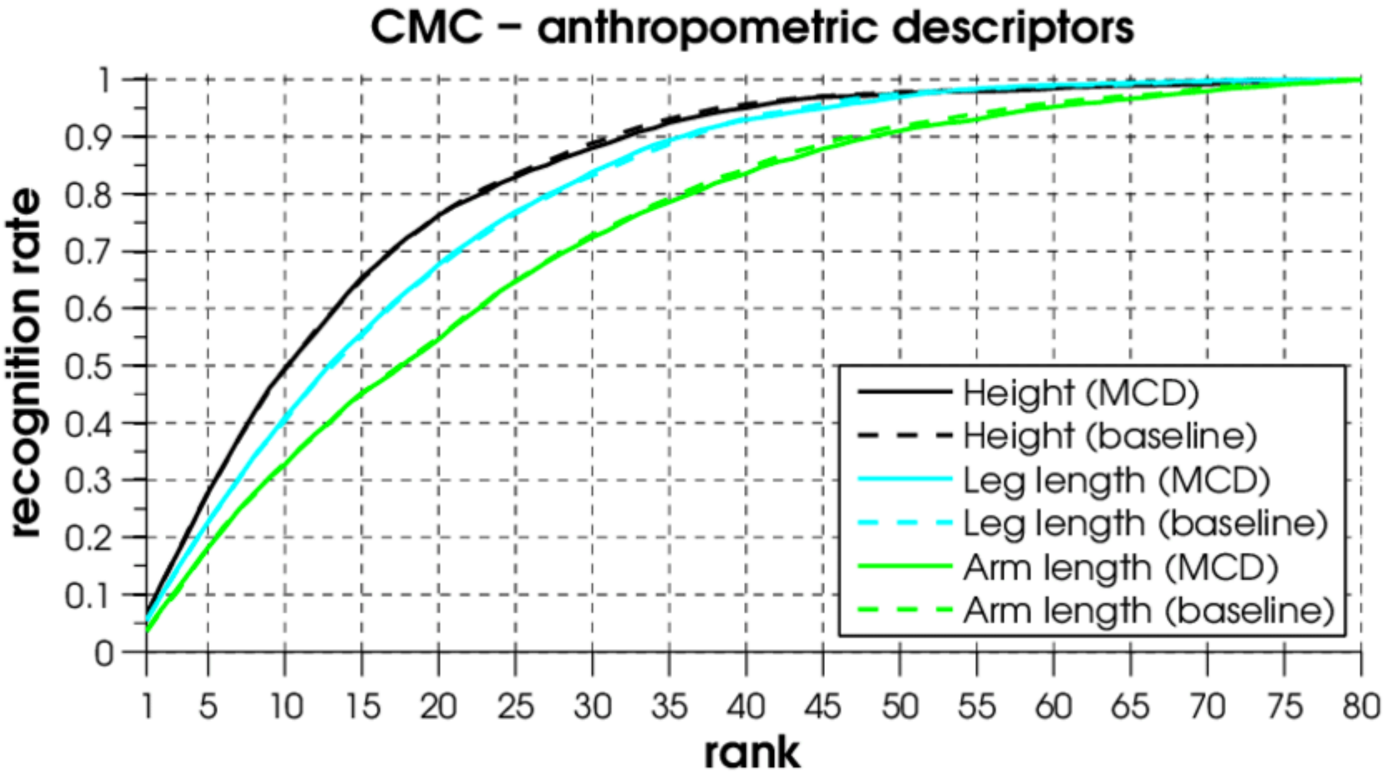
Performance Evaluation

- Receiver operating characteristics (ROC) curve
  - For authentication/verification
  - False positive rate vs true positive rate
- Detection error-tradeoff (DET) curve
  - False positive rate vs false negative rate
- Cumulative match curve (CMC)
  - Rank vs identification accuracy

ROC Curve



CMC Curve


CMC – anthropometric descriptors
recognition rate
rank
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
Height (MCD)
Height (baseline)
Leg length (MCD)
Leg length (baseline)
Arm length (MCD)
Arm length (baseline)