Pattern Recognition

EE468 (EEE&EIE), EE9SO29, EE9CS729

Tae-Kyun Kim
Senior Lecturer
http://www.iis.ee.ic.ac.uk/ComputerVision/

Lecture Schedules

9 week lectures (in autumn term)
- Lectures are in week 2-3, 5-11
- Every Tuesday, 4-6pm (2 hours)
- Room 509A EEE

1 week computer lab
- 19 Oct (Thu) in week 3
- 4-5pm, Room 408
- 5-6pm, Room 305 (computer lab)

100% coursework
- Computer programming based (Matlab, other tools)
- 2 courseworks (both due the end of lectures or the similar)
- (refer to those of PR in 2016)

Course homepage:
- https://intranet.ee.ic.ac.uk/electricalengineering/eecourses_t4/course_content.asp?c=EE4-68&s=T4
- https://bb.imperial.ac.uk

Lecture Schedules

Lecturer: Dr Tae-Kyun Kim

- Office: EEE 1017
- http://www.iis.ee.ic.ac.uk/ComputerVision/
- Week 2-3, Week 5-6

GTA : Guillermo Garcia-Hernando

- Email: g.garcia-hernando@imperial.ac.uk
- Office: EEE 1008d
- http://www.iis.ee.ic.ac.uk/ComputerVision/
- Computer lab in Week 3

Lecturer: Dr Krystian Mikolajczyk

- Office: EEE 1015
- http://www.imperial.ac.uk/people/k.mikolajczyk
- Week 7-11

Course Aims

- This course aims to introduce the concepts, basic formulations and applications of pattern recognition.
- This module is in close relations to: Machine Learning for Computer Vision EE462/EE9SO25/EE9CS728 (spring term).

- The module studies, given feature representation in a vector form, the concept of machine perception and decision surfaces, and metrics/distances, model fitting, as basic tools to process and classify data.
- Deep Convolutional Neural Network is taught as the-state-of-the-art PR method.

- The learnt topics are illustrated with few applications: face recognition, or machine learning repository data, handwritten digit recognition, etc.

Lecture Syllabuses

Topics

- Machine perception, Feature extraction, Subspace learning, PCA, Nearest neighbour classification
- Linear discriminant functions, Discriminant analysis, Bayes decision theory
- Decision hyperplane, Maximum margin, Linear classifier, Kernel trick, Nonlinear classifier
- Bagging and boosting, Resampling for estimation, Classifier combination, Estimation of misclassification
- Metrics/distances: Mahalanobis, Template matching, Hough transform, RANSAC, Clustering algorithms, Kmeans, Agglomerative, EM, Meanshift
- Deep neural networks: Perceptron concept, widely used DL architectures: RNN, LSTM, CNN, their component layers, parameters and optimisation by backpropagation.
- Practical sessions: Pattern recognition using Face dataset or ML repository (by Matlab and/or other tools)

Backgrounds

The module is coursework-based and the coursework requires Matlab programming.

The lectures require a background on:

- Linear algebra (EE310)
  - Orthogonal/orthonormal vectors
  - Basis vectors/Subspaces
- Optimisation (EE429)
  - Gradient method
  - Lagrange multipliers
- Matrix and vector derivatives

*Appendix A: Mathematical Foundations, R.Duda, P.Hart, D.Stork, Pattern Classifcation (Second Edition), JOHN WILEY & SONS, Inc. 2001.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.320.4607&rep=rep1&type=pdf
http://cns-classes.bu.edu/cn550/Readings/duda-etal-00.pdf

This module is closely related to

- EEE courses: EE462 Machine Learning for Computer Vision, Intro to Machine Learning
- Computing courses: 316 Computer Vision, 395 Machine Learning, 333 Robotics, 495 Advanced Statistical Machine Learning and Pattern Recognition

Introduction to Pattern Recognition

Tae-Kyun Kim
Senior Lecturer
http://www.iis.ee.ic.ac.uk/ComputerVision/

What is Statistical Pattern Processing

What is a Pattern?

- A **pattern**, apart from the term's use to mean "Template", is a discernible regularity in the world or in a manmade design. As such, the elements of a pattern repeat in a predictable manner.
- A pattern is an abstract object, such as a set of measurements describing a physical object.
- These patterns can represent many different types of object (speech/image/text etc).

The main area of Statistical Pattern Processing discussed in this course is **classification** of patterns into different classes.

- Concepts, Theory, Algorithms, Systems to put Patterns into Categories
- Classification of High-dimensional, Complex, or Noisy Data
- Relate Perceived Pattern to Previously Perceived Patterns

What is Statistical Pattern Processing

A key issue in all pattern recognition systems is variability.

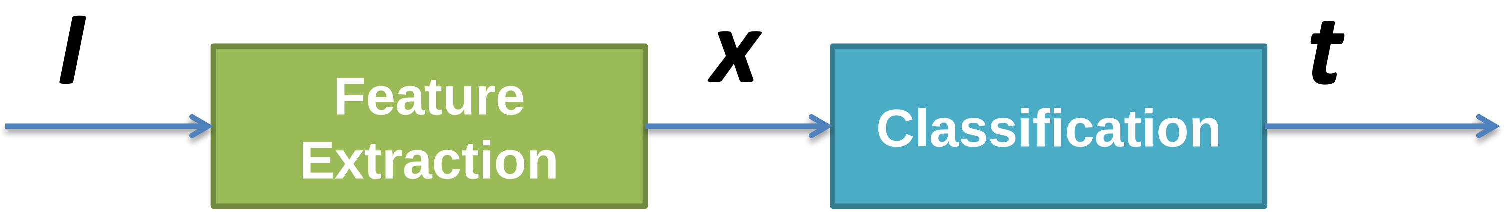
- Patterns arise (often from natural sources) that contain variations.
- Are the variations systematic (and can be used to distinguish between classes?
- Or are they noise?
- The variability of classes can be more explicitly approached by using probabilistic generative (cf. discriminative) modelling of pattern variations.

The standard model for pattern recognition divides the problem into two parts:

- Feature extraction
- Classification

Basic Model

- Initial feature extraction produces a vector of features that contain all the information for subsequent processing (such as classification).
- Ideally, for classification, only the features that contain discriminatory information are used.
- Often features to measure are determined by an “expert”, although techniques exist for choosing suitable features.
- The classifier processes the vector of features and chooses a particular class.
- Normally the classifier is “trained” using a set of data for which there are labelled pairs of feature vectors / class identifiers available.

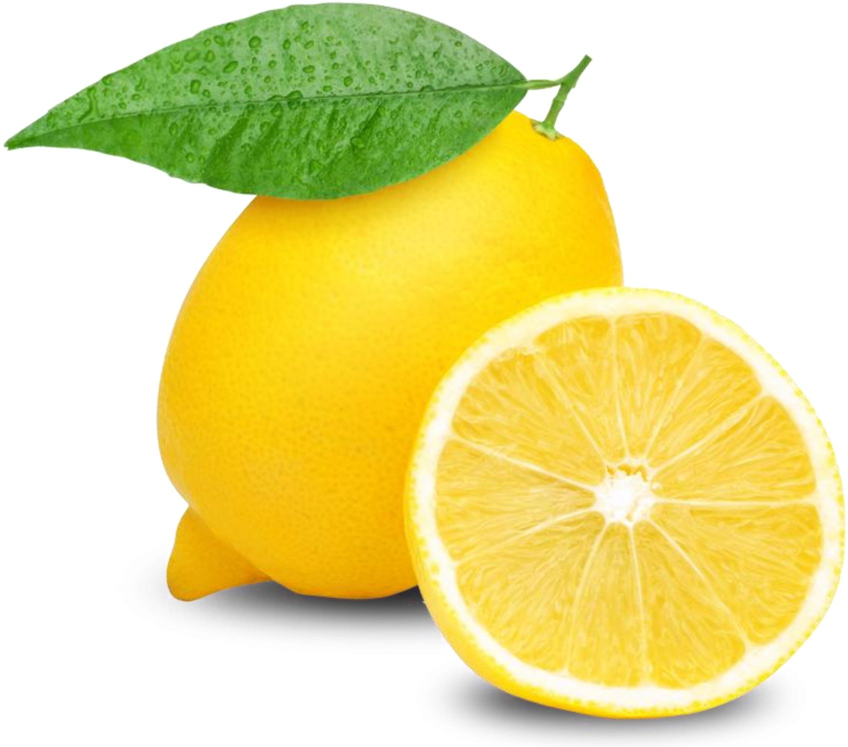


Pattern e.g. images, speech signals

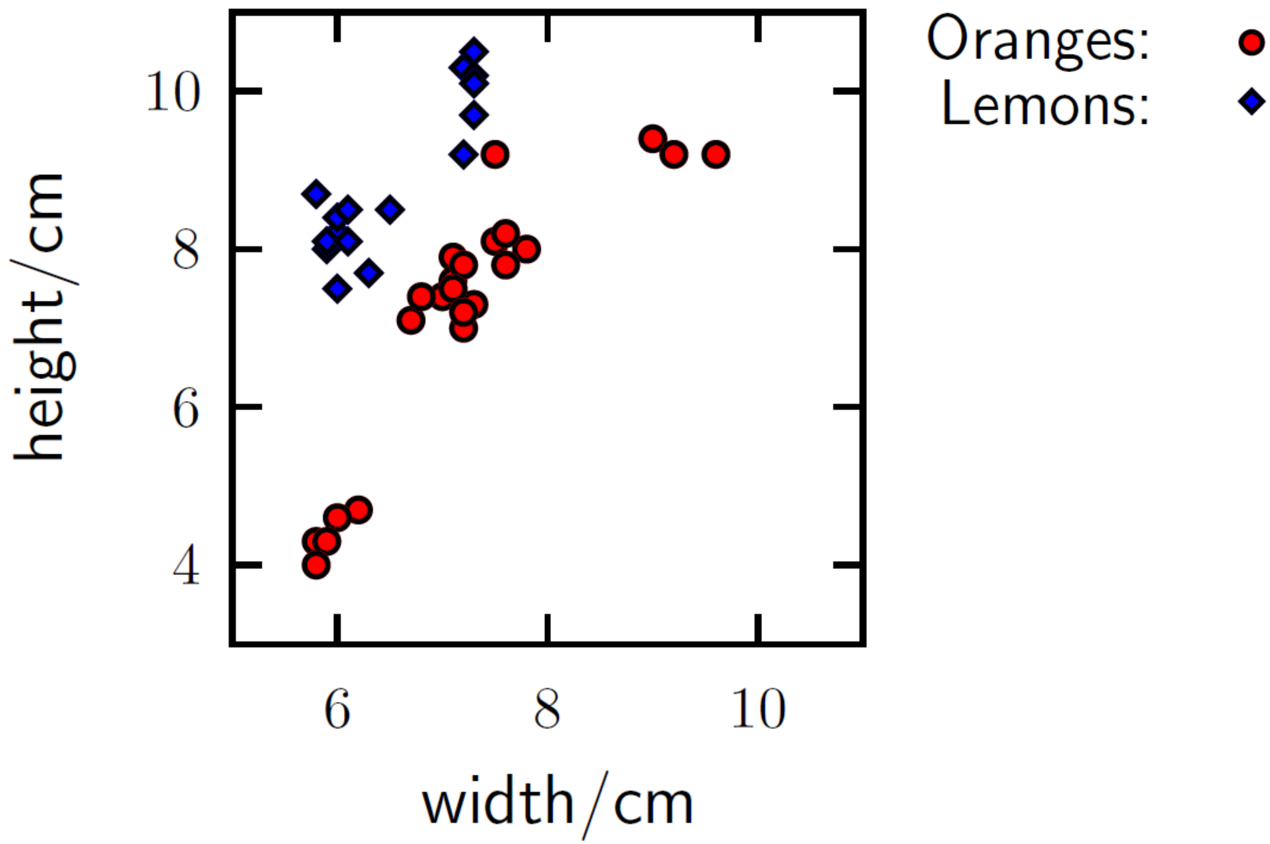
Feature vector e.g. principal/spectral components

Symbol e.g. object classes/categorie s, words

Example Problem:
Oranges and Lemons

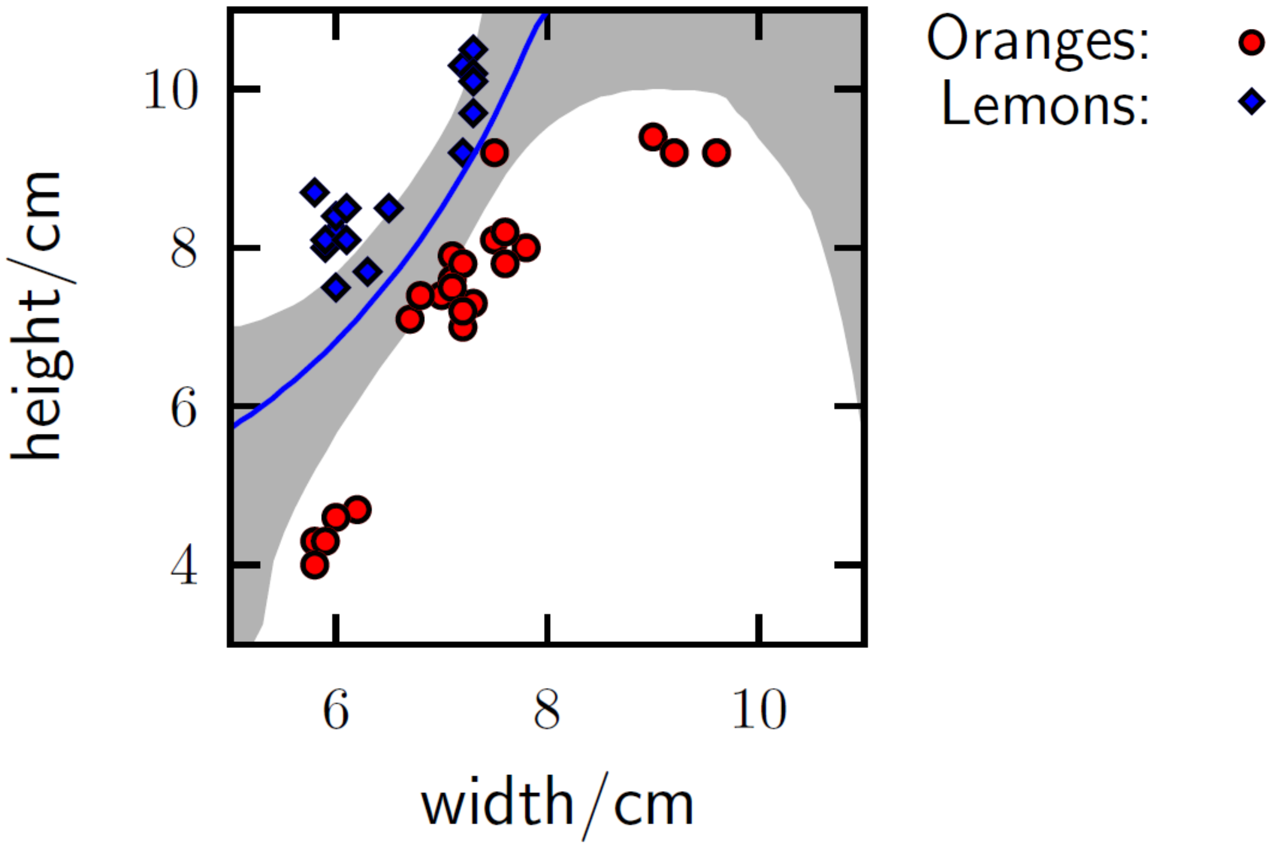


Example Problem: Oranges and Lemons



A two-dimensional space

Example Problem: Oranges and Lemons

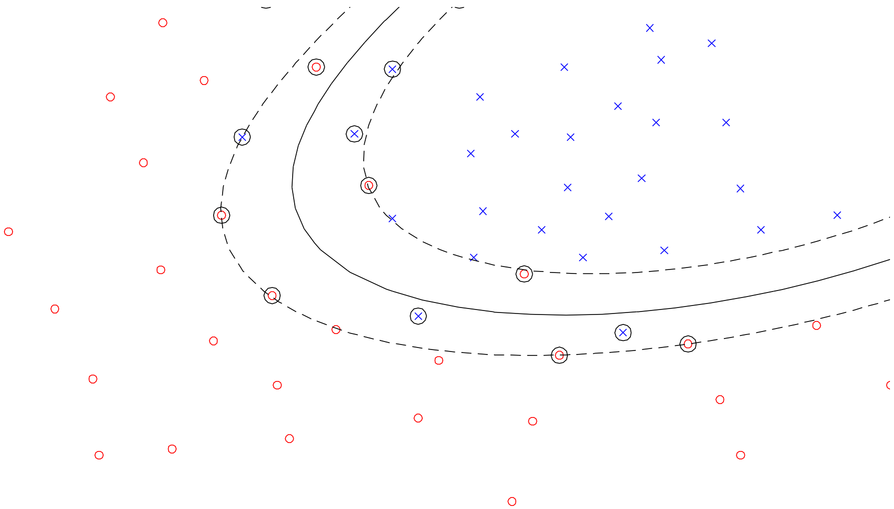


A two-dimensional space

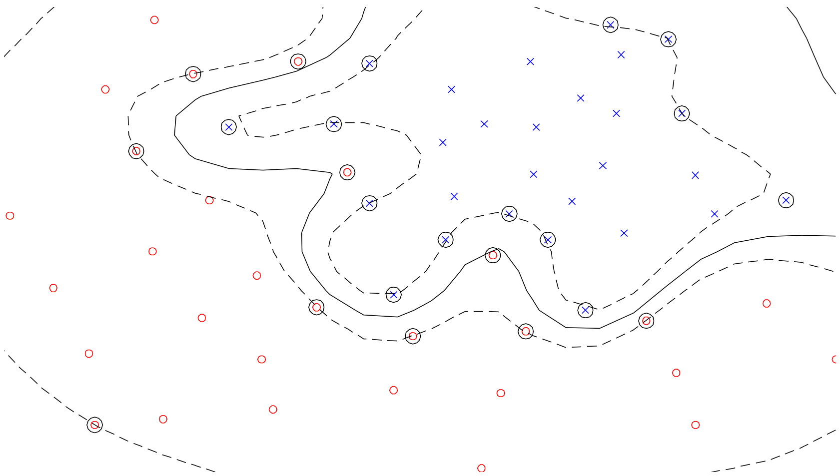
Generalization

Simple models have better generalization to unseen data, complex models can be overfitted to training data.

Linear models, or a combined set of linear models work well in practice.

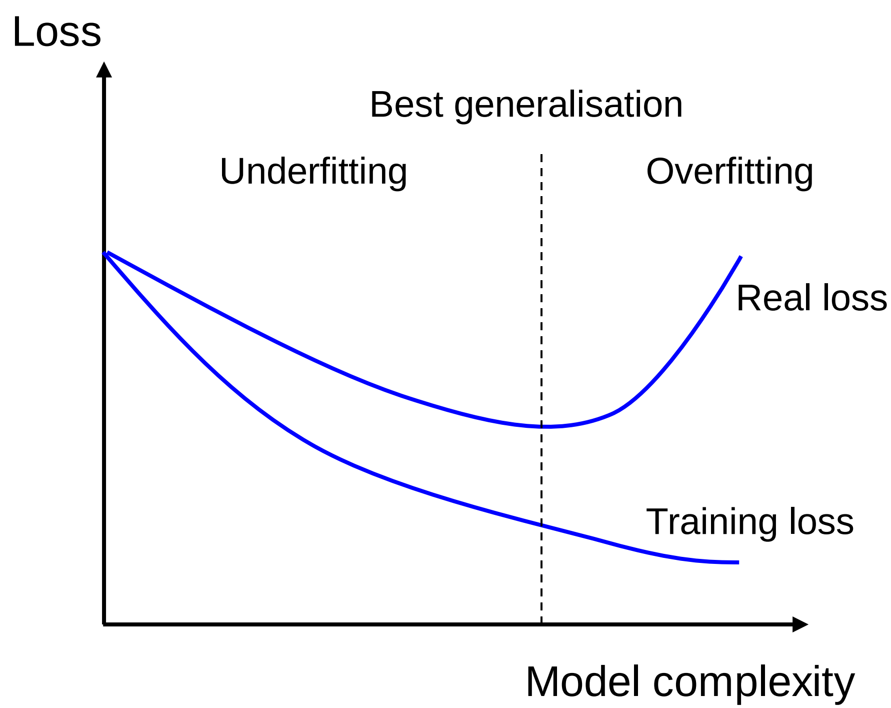


Overfit



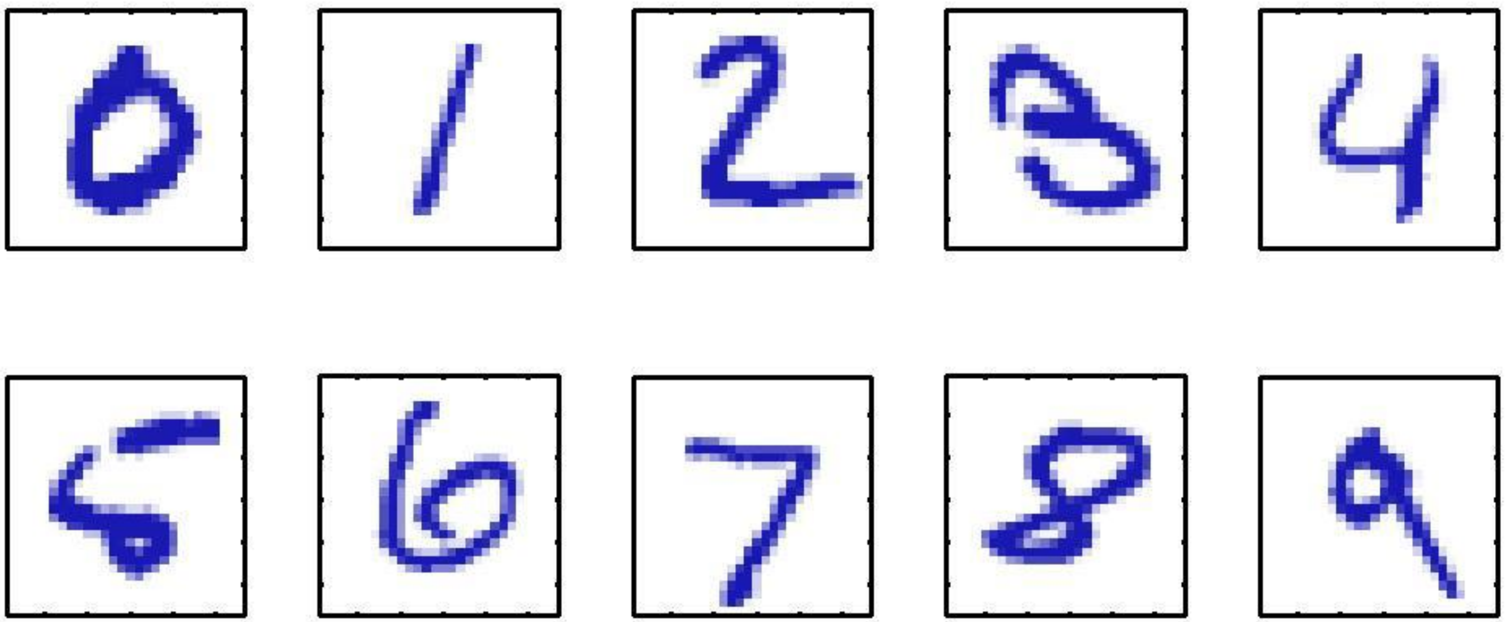
Overfitting

- As complexity increases, the model overfits the data:
  - Training loss (classification error of training data) decreases.
  - Real loss (classification error of testing data) increases.
- We need to penalize model complexity = to generalise.



Example Problem: Handwritten Digit Recognition

– Wide variability of same numeral
– Handcrafted rules/features result in a large number of rules and exceptions
– Better to learn features from a sized training/example set



Example Problem: Handwritten Digit Recognition

- Consider an example of digit images that undergo a random displacement and rotation.
- The images have the size of 100 x 100 pixel values, but the degree of freedom of variability across images is only three: vertical, horizontal translations and rotations.
- The data points live on *a feature space whose intrinsic dimensionality* is three.
- The translation and rotation parameters are continuous *feature variables*. We only observe the image vectors.