Pattern Recognition (PR) Course
EQ2340
KTH

Year 2017, Period 1

Book: Pattern Recognition Compendium (By Arne Leijon and Gustav Henter)

This document contains the pointers that we cover in class

Lecture 1

Introduction:

1. Why Pattern Recognition? (a) Several problems: Visual recognition, Speech recognition, Hand writing recognition, Detecting radio signal corrupted in noise, in fact, we can give hundreds of examples. (b) Possibly human is the best in Pattern Recognition – opportunities to use PR theory and human brain function theory in conjunction, to understand each other.
2. Basic system architecture (block diagram) of a PR system – why feature extraction? Curse of dimensionality.
3. What fundmanetal questions we cover in the course? See Table 1. Some examples
4. Question to ask? Can we use an optimal single decision (binary decision) classifier repeatedly for the task of multiple classification?
5. What we will cover in the course? See the last paragraph of page 2 in compendium.

Chapter 1: Classification and Probability

1. What we are going to do? Desgning a simple two-category (binary) signal classification system.
2. We begin with a simple Example. Example 1.1 and discussion. Figure 1.1
3. Decision function, eq. 1.1, Decision region, eq. 1.2, Discriminant function, eq. 1.3, 1.4, linear discriminant, eq. 1.5
4. Gaussian feature vector, Figure 1.2, Figure 1.3
5. Two independent features, eq. 1.6-1.10, General correlated features, eq. 1.11-1.16,
6. Decorrelation of Gaussian source – Coordinate transformation and eigen vectors, eq. 1.17-1.21
7. Gaussian mixture model, why GMM? eq. 1.25-1.27, Figure 1.5 and Figure 1.6, Example 1.1
8. Summary of the Chapter 1 (page no 15 of the book).

Lecture 2

Chapter 2: Conditional Probability and Bayes Rule

1. Joint density, conditional density and Bayes rule
2. Sum rule and product rule

Chapter 3: Bayesian Pattern Classification

1. Figure 3.1
2. Problem setup, three main points (bullet points in page no 29)
3. A priori source category distribution, eq. 3.1,
4. Feature vector distribution

5. A posteriori form, eq. 3.3-3.7
6. Bayes minimum risk decision rule, Loss matrix, eq. 3.8; conditional expected loss (conditional risk), eq. 3.9; Bayes minimum risk decision rule, eq. 3.10; eq. 3.11
7. Special case: minimum error rate, eq. 3.12-3.13, MAP rule, eq. 3.14-3.15
8. Special case: Equal a-priori probability, ML rule, eq. 3.16

Lecture 3

Chapter 3:

1. Discriminant functions, General decision rule, discriminant functions, eq. 3.17, 3.18; monotonically increasing function (log function), eq. 3.19-3.21.
2. Decision regions and error probability, decision regions, eq. 3.24, Figure 3.2, 3.3, eq. 3.25, 3.26;
3. Simple example: Scalar Gaussian feature, Problem statement, Solution approach (in five main steps). Let us have a rough discussion following the steps in book, but the teacher will not write all steps one-by-one on the board.
4. Summary of the chapter 3.

Assignment for students: Solve following the book approach – (a) Example of scalar Gaussian feature (section 3.5) , (b) Example of independent Gaussian feature (section 3.6), also the multivariate Gaussian cases (section 3.7)

Note an important point: How the discrimiant functions become linear.

Lecture 4

Chapter 5: Hidden Markov Models (HMM) for sequence classification

1. Why HMM? Sequnce of decisions, example of speech recognition (window wise feature extraction using stationarity), equantion 5.1-5.3, too many combinations for possible state sequences, MAP rule and equation 5.4, Three reasons of using HMM
2. HMM definiton, figure 5.1, Three key factors for HMM simplicity, HMM system explanation (clearly state the parameters), Initial state probability (eq. 5.5), Transition probability matrix (eq. 5.6), Output probability distribution (eq. 5.7), Continuous valued observations – use of GMM, eq. 5.8, tied observation (eq. 5.10).
3. HMM representations, eq. 5.14, State graph, State-time graph and Bayesian network graph (Figure 5.2, 5.3, 5.4),.
4. HMM structures, only few pointers – Left-right HMM, Figure 5.7, Irreducible.
5. Probability of an observed sequence – the Forward algorithm (We discuss and prove steps in detail).

Assignment for students: Rework the derivation of forward algorithm before coming to next lecture class.

Lecture 5

Chapter 5: Hidden Markov Models (HMM) for sequence classification

1. We finish the Forward algorithm
1. And Start the Backward Algorithm (section 5.5) - Introduction.

If possible, we will start section 5.6 : Most probable state sequence – the Viterbi algorithm. (We discuss and prove all steps in detail).

Lecture 6

Chapter 5: Hidden Markov Models (HMM) for sequence classification

1. The Backward Algorithm (section 5.5) - (We discuss and prove all steps in detail). Discuss the importance of backward algorithm. We will cover section 5.5.1 and section 5.5.2. Introducing all relevant variables, the main equations are: 5.56-5.70.
2. Most probable state sequence – the Viterbi algorithm (section 5.6); decoding problem, the concept of dynamic programming, Viterbi decoding in step-by-step (eq. 5.71-5.83), Viterbi probability vector, backtracking matrix, recursive formulation of eq. 5.80, backtracking.
3. Summary of Chapter 5.

Lecture 7

Chapter 6: Hidden Markov Model Training

Before discussion, write down the following equations on a corner of the board.

1. Total conditional state probability - eq. 5.63, Forward variable - eq. 5.30, Forward scale factor – eq. 5.33, Scaled forward variable – eq. 5.38, Total likelihood – eq. 5.36, Backward variable – eq. 5.59, scaled backward variable – eq. 5.61

Discussion agenda:

1. Motivation of this chapter, training data – test data, Baum Welch algorithm, connection with EM algorithm
2. The Baum-Welch Algorithm, Problem statement (eq. 6.1), comment on optimality; Iterations of Baum-Welch algo: Initialization, repeat, terminate and trim (eq. 6.2-6.3); crucial point on how to choose the parameters – we proceed intuitively, but not by a formal theoretical argument/proof; Comment on Practical Issues: Several instances (or words), finite and infinite HMM cases, the HMM should not model the discontinuity between two uttered training instances;
3. Data representations, eq. 6.4-6.5, parameter update formulation by eq. 6.6;
4. Updating the initial proability vector, eq. 6.7-6.10; Updating the transition probability matrix, eq. 6.11-6.19;
5. Updating the observation probability matrix, eq. 6.20-6.22
6. Practical issues: Follow the book and discuss few main pointers – feature extraction, model size or complexity, Initializing a left-right HMM, Initializing an ergodic HMM, Termination of HMM training, Adjustment after training.
7. Summary of the chapter

Lecture 8

Discussion Agenda:

Chapter 6: Hidden Markov Model Training

1. Practical issues: Follow the book and discuss few main pointers – feature extraction, model size or complexity, Initializing a left-right HMM, Initializing an ergodic HMM, Termination of HMM training, Adjustment after training.
2. Summary of chapter 6

Chapter 7: Expectation Maximization

1. Baum-Welch algo is an instance of EM, Why EM (latent / hidden variable concept).
2. About the fundamental "trick" in EM – somehow get an estimation / idea about the hidden ones
3. EM: Intuitive Introduction (chapter 7.1) – Problem stattement (eq. 7.1), then we discuss eq. 7.2-7.5, A difficulty of random and deterministic cost, Introduction of Expectation sense, Help function – eq. 7.6, itaerative step – eq. 7.7
4. EM algorithm steps – follow the book. At this point, we may also follow Bishop book and write the steps of General EM algo using proper notations – according to the notations in the text book.
5. EM algorithm proof (section 7.3), Theorem 7.1 (eq. 7.11-7.12), Proof of the theorem, Jensen's inequality (Show Page no 56 of Bishop Book).
6. Example 7.1 (Missing data), eq. 7.13-7.25
7. Training of GMM (section 7.5), eq. 7.26-7.27, Objective,

Lecture 9

Chapter 7: Expectation Maximization

Before discussion, write down the following equations on a corner of the board.
1. eq. 5.57, eq. 6.14

Discussion agenda:

1. Example 7.1 (Missing data), eq. 7.13-7.25
2. Training of GMM (section 7.5), eq. 7.26-7.46,
3. HMM Training (section 7.6), eq. 7.47-7.70
4. Summary of Chapter 7

Lecture 10

Chapter 7: Expectation Maximization

We finish discussion on chapter 7. Then start Chapter 8.

Chapter 8: Bayesian learning

New Discussion agenda:

1. ML approach – main four steps to perform ML, the disadvantages of ML, so comes Bayesian learning
2. Bayesian learning – Introduction (section 8.1), state four main advantages, the single crucial theoretical concept – make everything probabilistic, some downside.
3. Example 8.1 (word recognition) – in detail, we consider all the steps and all equations 8.1-8.11. (show Beta density picture in page no 686 of Bishop book)
4. In Example 8.1, we discuss on Beta distribution as the normal choice of unknown parameter, and then formulate the notion of conjugate prior.
5. Bayesian learning procedure (section 8.2) – there are six steps and we discuss the steps step-by-step.