Pointers for EQ2415 Machine Learning and Data Science Course

Saikat Chatterjee

Abstract

Index Terms

***

Date: August 26, 2019

I. Projects

Object recognition problem: In computer vision, object recognition is a standard machine learning problem with high impact. Throughout the course, we will do object recognition in the scope of projects using any algorithms. Please use the following databases and try several algorithms one-by-one as you learn them in the course. Finally, critically compare all the algorithms from level of complexity, implementation issues and algorithm fundamentals. The databases are mentioned in Table I.

For more details, please read Experiment section of the article in Arxiv. "SSFN: Self Size-estimating Feed-forward Network and Low Complexity Design", Saikat Chatterjee, Alireza M. Javid, Mostafa Sadeghi, Partha P. Mitra, Mikael Skoglund. Search in Google and you will get the Arxiv link for the paper. You may even read the whole paper for better understanding.

II. Starting with self-study

Please start with self-study from Bishop's book. Please note the points below.

1) Bishop starts with a simple polynomial curve fitting example. Study material: Bishop's Book, chapter 1.1
2) Revisit the curve fitting problem in a Bayesian setup. The reason is understand relation between regularized least-squares problem and Bayesian problem. Study material: Bishop's Book, chapter 1.2.5

III. Linear Inference

The general regression setup is $\hat{\mathbf{t}} = \mathbf{A}\boldsymbol{\phi}(\mathbf{x})$. We start with linear regression. The data-and-target-pair is $(\mathbf{x}, \mathbf{t})$. Using a linear system, we predict target as

$$\hat{\mathbf{t}} = \mathbf{A}\mathbf{x}. \tag{1}$$

1) How to solve $\mathbf{A}$ matrix by formulating a cost function?
2) What is the role of regularization?
3) Least-squares and minimum-mean-square-error formulation.
4) Why we like linear system from signal-and-systems based analysis?
5) When a linear system is stable? Noise robustness and reliable.

Study material: Bishop's Book, chapters 3.1, 3.1.1, 3.1.4, 3.1.5.

Project: Please implement object recognition using linear inference setup. Establish a regularized cost function in constrained and Lagrangian formats. Using ℓ_2-norm, the solution is a least-squares solution, also known as Tikonov regularization. Find optimal regularization constant using multi-fold cross validation. Tabulate the results.

IV. Kernel substitution

A. *Kernel substitution for a linear inference*

Study material: Bishop's Book, chapter 6.1 and 6.2

We predict the target as

$$\hat{\mathbf{t}} = \mathbf{A}\boldsymbol{\phi}(\mathbf{x}). \tag{2}$$

Then we show kernel formulation – using inner product formulation without actually knowing the exact form of the feature vector.

Project: Please implement object recognition using kernel substitution approach. Establish a regularized cost function in constrained and Lagrangian formats. Use ℓ_2-norm and then form a dual. Then show the standard expression of kernel predictor. Use Gaussian kernel. Find optimal regularization constant using multi-fold cross validation. Tabulate the results.

TABLE I: Databases for multi-class classification

Database	# of train data	# of test data	Input dimension (P)	# of classes (Q)	Random Partition
Vowel	528	462	10	11	No
Extended YaleB	1600	800	504	38	Yes
AR	1800	800	540	100	Yes
Satimage	4435	2000	36	6	No
Scene15	3000	1400	3000	15	Yes
Caltech101	6000	3000	3000	102	Yes

B. Support vector machine

Then we are supposed to read about support vector machine (SVM) by our own.
Study material: Bishop's Book, chapter 7.1 and 7.2

C. Gaussian Process

Please read chapter 6.4.1 and 6.4.2 from Bishop's book. Equations 6.64 - 6.67 are important and please read carefully upto the proof of these equations.

V. Neural Networks

A. Challenges in design of large neural networks

Study material: Slides from Saikat

Project: Please implement object recognition using a single-layer extreme learning machine (ELM). Establish a regularized cost function in constrained and Lagrangian formats for learning of output matrix of ELM. Use ℓ_2-norm and show that the relevant problem has the same solution as Tikonov regularization. Find optimal regularization constant using multi-fold cross validation. Tabulate the results. Then plot accuracy versus number of nodes in the ELM.

Then implement a three layer ELM. Think how to choose the size of the 3-layer ELM.

Reading Project: Read the following article. "SSFN: Self Size-estimating Feed-forward Network and Low Complexity Design", by Saikat Chatterjee, Alireza M. Javid, Mostafa Sadeghi, Partha P. Mitra, Mikael Skoglund.
https://arxiv.org/pdf/1905.07111.pdf
Then, implement back propagation algorithm for the Self Size-estimating Feed-forward Network (SSFN).

B. RBM - Restricted Boltzman Machine

Study material: From Wikipedia and relate with Graphical Models taught by Ragnar Thobaben

VI. Sparse Representation

Following Saikat's slides, we have discussed following issues.

1) Underdetermined linear model for sparse representation.
2) Regularization using ℓ_1-norm
3) P_0 and P_1 problems.
4) Basis pursuit and orthogonal matching pursuit algorithms
5) Stability of algorithms using mutual coherence and restricted isometry properties
6) Pattern recognition using sparse representation.
7) Dictionary learning - a Factorized Matrix Learning Problem.
8) Lebel-consistent dictionary learning for pattern recognition
9) Proof for Orthogonal Matching Pursuit (OMP) algorithm on its recovery performance.
10) Connection with OMP and Neural Network.
11) Non-negative Matrix Factorization - How similar the problem is with dictionary learning.

Reference:

1) Sparse and redundant representations: from theory to applications in signal and image processing, M. Elad.
2) "Robust Face Recognition via Sparse Representation", IEEE Trans Pattern Analysis and Machine Intelligence
3) "Label-consistent K-SVD: Learning a discriminative dictionary for recognition", IEEE Trans Pattern Analysis and Machine Intelligence

Project: Let us assume $\mathbf{x} \in \mathbb{R}^m$ is a sparse vector with sparsity level $k \triangleq \|\mathbf{x}\|_0 << m$. For example, you can assume k is close 5% of m. This is motivated by many image signals where $5 - 10\%$ of Wavelet or DCT coefficients hold almost all energy of the image. Then, generate a random matrix instance $\mathbf{A} \in \mathbb{R}^{n \times m}$ and create observation $\mathbf{b} = \mathbf{A}\mathbf{x}$. Then reconstruct $\mathbf{x}$ from observation $\mathbf{b}$. Plot the reconstruction measure using normalized-mean-square-error (NMSE) versus the value of $\frac{n}{m}$. That means reconstruction quality versus number of observations. Use Basis Pursuit and Orthogonal Matching Pursuit algorithms for reconstruction. Do you see any phase transition in your results? Here phase transition means that you see sudden change from bad reconstruction quality to very good reconstruction quality when you increase number of observations.