# The Final Project of "introduction to Statistical Learning and Machine Learning"

Yanwei Fu

December 25, 2017

**Abstract**

(1) This is the final project of our course. The project is released on Dec 25th, 2017. The deadline is 5:00pm, Feb. 7th, 2018. Please send the report to 582195191@qq.com, or 17210240267@fudan.edu.cn. The late submission is also acceptable; however, you will be penalized 10% of total scores for EVERY TWO DAYS' delay (by 5:00pm of that day).

(2) You will get 3-4 papers soon after for you to review. That is, you will review anonymously (double blind) 3 papers submitted by your colleagues. The reviews are due back on Feb. 10th at 5pm (tentative). Each review should be about 1 page. It must assign marks between 1 and 10 to each paper, as suggested by NIPS instructions. For submission of your reviews, email me a PDF file of 3-4 pages only. The title of each page should be the paper title of the paper under review with your score in brackets. In general, we will try to assign each paper with at least 5 reviewers. If you are failed to submit the reviews on time for one paper-review you assigned, you will get 1% penalty of your total scores for EACH DAY's delay on this project.

(4) Note that if you are not satisfied with the initial report, the updated report will also be acceptable given the necessary score penalty of late submission.

(5) OK! That's all. Please let me know if you have any additional doubts of this project. Enjoy!

Note that:

(a) If the size of training instances are too large, you may want to apply some sampling techniques to extract a small portion of training data.

(b) If you think the dimension of features is too high, you may also want to use some techniques to do feature dimension reduction, such as PCA, KPCA, ISOMAP.

(c) The referring papers are listed as an introduction to the context of problems. It's not necessarily to exactly implement these papers, which actually is not an easy task.

**(d) For all the projects, we DO care about the performance on each dataset with the correct evaluation settings.**

Merry Christmas!

# 1 Introduction

## 1.1 Collaboration Policy

You are allowed to work in a group with at most two collaborators. you will be graded on the creativity of your solutions, and the clarity with which you are able to explain them. If your solution does not live up to your expectations, then you should explain why and provide some ideas on how to improve it. You are free to use any third-party ideas or codes that you wish as long as it is publicly available. You must provide references to any work that is not your own in the write-up.

### 1.2 Writing Policy

Final project (20%) is finished by one team. Each team should have up to 3 students; and will solve a real-world Big-Data problem. The final report should be written in English. The main components of the report will cover

1. Introduction to background and potential applications (2%);
2. Review of the state-of-the-art (3%);
3. Algorithms and critical codes in a nutshell (10%);
4. Experimental analysis and discussion of proposed methodology (5%).

Please refer to our latex example: http://yanweifu.github.io/courses/chap5_svm%281%29/IEEE_TAC_2016.zip.

### 1.3 Submitting Policy

The paper must be in NIPS format (downloadable from [1]) and it must be double-blind. That is, you are not allowed to write your name on it etc. For more info, please read: NIPS reviewieng and double blind policy.

Package your code and a copy of the write-up pdf document into a zip or tar.gz file called finalProject-*your-student-id1_student-id2_student-id3.[zip|tar.gz]. Also include functions and scripts that you had used. To submit the report, email the pdf file to 582195191@qq.com, or 17210240267@fudan.edu.cn. In the submission email, you should well explain the authours and co-workers of this project.

### 1.4 Evaluation of Final Projects

You will get 3-4 papers soon after for you to review. That is, you will review anonymously (double blind) 3 papers submitted by your colleagues. The reviews are due back on Feb. 22nd at 5pm. Each review should be about 1 page. It must assign marks between 1 and 10 to each paper, as suggested by NIPS instructions. For submission of your reviews, email me a PDF file of 3-4 pages only. The title of each page should be the paper title of the paper under review with your score in brackets. e.g. "Least Squares for Energy Prediction (7)". The rest of each page (one page per review) should be a text evaluation of the work you are reviewing. It should be based on the following NIPS criteria:

**Overview:** you should briefly summarize the main content of this paper, as well as the Pros and Cons (advantages and disadvantage) in general. This part aims at showing that you had read and at least understand this paper.

**Quality:** Is the paper technically sound? Are claims well-supported by theoretical analysis or experimental results? Is this a complete piece of work, or merely a position paper? Are the authors careful (and honest) about evaluating both the strengths and weaknesses of the work?

**Clarity:** Is the paper clearly written? Is it well-organized? (If not, feel free to make suggestions to improve the manuscript.) Does it adequately inform the reader? (A superbly written paper provides enough information for the expert reader to reproduce its results.)

**Originality:** Are the problems or approaches new? Is this a novel combination of familiar techniques? Is it clear how this work differs from previous contributions? Is related work adequately referenced?

[1] https://nips.cc/Conferences/2016/PaperInformation/StyleFiles

**Significance:** Are the results important? Are other people (practitioners or researchers) likely to use these ideas or build on them? Does the paper address a difficult problem in a better way than previous research? Does it advance the state of the art in a demonstrable way? Does it provide unique data, unique conclusions on existing data, or a unique theoretical or pragmatic approach?

You project mark will mostly be based on the scores you get from your colleagues. I will simply play the role of chair and calibrate the scores to make sure there is no bias. I will also control 20-40% of the mark and this will be based on the quality of the reviews.

#### 1.4.1 Minimum Requirements

For all the projects listed below, in general you should devise your own machine learning algorithms which target at each specific problem of each project. You should compare with the machine learning algorithms taught in this course/mini-projects, which, include but not limited to, *linear regression/classification, K-NN/NN, logistic regression, linear/RBF kernel SVM, Neural network* as well as *tree-based methods*. Thus, the minimum requirements, as you can image, just apply and compare with these methods; and explain the advantage and disadvantage of using these methods for the project problem. Note that, your algorithms can be derived from one of these machine learning algorithms; and feel free to use any machine learning package you like.

## 2 Potential Projects

### 2.1 One-shot Learner

#### 2.1.1 Introduction to this project

The success of recent machine learning (especially the deep learning) greatly relies on the training process on hundreds or thousands of labelled training instances of each class. However in practice, it might be extremely expensive or infeasible to obtain many labelled data, e.g. for objects in dangerous environment with limited access. On the other hand, human can recognize an object category easily with only a few shots of training examples. Inspired by such an ability of humans, one-shot learning aims at building classifiers from a few or even a single example. One-shot learning means that only one or very few training examples are available to train classifiers. Of course, the major obstacle of learning good classifiers in one-shot learning setting is the lack of enough training data.

Existing one-shot learning approaches can be divided into two groups: the direct supervised learning based approaches and the transfer learning based approaches.

- Direct Supervised Learning-based Approaches: Early approaches do not assume that there exist a set of auxiliary classes which are related and/or have ample training samples whereby transferable knowledge can be extracted to compensate for the lack of training samples. Instead, the target classes are used to trained a standard classifier using supervised learning. The simplest method is to employ nonparametric models such as kNN which are not restricted by the number of training samples. However, without any learning, the distance metric used for kNN is often inaccurate. To overcome this problem, metric embedding can be learned and then used for kNN classification. Other approaches attempt to synthesize more training samples to augment the small training dataset [15]. However, without knowledge transfer from other classes, the performance of direct supervised learning based approaches is typically weak. Importantly, these models cannot meet the requirement of lifelong learning, that is, when new unseen classes are added, the learned classifier should still be able to recognize the seen existing classes.

- Transfer Learning-based One-shot Recognition: This category of approaches follow a similar setting to zero-shot learning, that is, they assume that an auxiliary set of training data from different classes exist. They explore the paradigm of learning to learn [20] or meta-learning [13] and aim to transfer knowledge from the auxiliary dataset to the target dataset with one or few examples per class. These approaches differ in (i) what knowledge is transferred and (ii) how the knowledge is represented. Specifically, the knowledge can be extracted and shared in the form of model prior in a generative model [4], features [10], and semantic attributes [8, 17, 19]. Many of these approaches take a similar strategy as the existing zero-shot learning approaches and transfer knowledge via a shared embedding space. Embedding space can typically be formulated using neural networks (*e.g.*, siamese network [14]), discriminative classifiers (*e.g.*, Support Vector Regressors (SVR) [3, 17]), or kernel embedding [10] methods. Particularly, one of most common embedding ways is semantic embedding which is normally explored by projecting the visual features and semantic entities into a common *new* space. Such projections can take various forms with corresponding loss functions, such as SJE [2], WSABIE [22], ALE [1], DeViSE [5], and CCA [7].

We use mini-imagenet dataset for this task. And the dataset of this project downloaded from our webpage[2].

As for the evaluation metrics, the basic experimental protocol is on "Transfer Learning-based One-shot Recognition", which should be strictly following Sec. 4.1.2 in [21]. USING WRONG EVALUATION METRICS WILL LEAD TO LOWER SCORE. Also note that, your work should be compared against several baselines. We do care the performance and accuracy reported.

You can also try the experimental protocol on "Direct Supervised Learning-based Approaches". Successful approaches on this aspect are very encouraging. Note that, still you need to compare several baselines.

#### 2.1.2 Submission and Evaluation

Note that, you should not copy any sentences from any paper. Remember the definition of plagiarism.

### 2.2 Zero-shot Learning

A major topic in this research area is building recognition models capable of recognizing novel visual categories that have no associated labelled training samples (i.e., zero-shot learning), few training examples (i.e. one-shot learning), and recognizing the visual categories under an 'open-set' setting where the testing instance could belong to either seen or unseen/novel categories. These problems can be solved under the setting of transfer learning. Typically, transfer learning emphasizes the transfer of knowledge across domains, tasks, and distributions that are similar but not the same. Transfer learning refers to the problem of applying the knowledge learned in one or more auxiliary tasks/domains/sources to develop an effective model for a target task/domain.

You need to do experiments on two datasets for zero-shot learning tasks. One dataset is USAA dataset which is for understanding event/activity happened in the video. The USAA dataset is well described and can be downloaded in USAA[3]. The low-level features for each video have been extracted. You may want to read our papers [6][8].

The another dataset is Animal with Attributes (AwA) dataset [16]. Note that AwA has two versions. Either one is good enough. The two versions can be downloaded from AwA1[4] or AwA2 [5]. The recent deep learning features as well as example codes of AwA dataset can be downloaded from AwA1[6] or Ugly-AwA [7]

[2] `http://www.sdspeople.fudan.edu.cn/fuyanwei/`
[3] `http://yanweifu.github.io/USAA/download/`
[4] `https://cvml.ist.ac.at/AwA/`
[5] `https://cvml.ist.ac.at/AwA2/`
[6] `https://github.com/pujols/zero-shot-learning`
[7] `https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/zero-shot-learning/zero-shot-learning-the-good-the-bad-and-the-ugly/`

As for the evaluation metrics

#### 2.2.1 Minimum requirements

On USAA dataset, you need to compare both supervised learning and zero-shot learning scenarios. Please refer to the experimental parts [8].

1, For supervised learning, we have pre-defined split and the testing data also come with the ground-truth. You can just try it. The dataset has the training and testing data, which are only for supervised learning.

2, As for zero-shot learning (ZSL), we use three splits for ZSL, and the zero-shot (testing classes) are [1,2,4,7], [1,6,7,8],[2,4,5,6]. For the details of the website: USAA[8]. Specifically,

> To repeat the Zero-shot learning results of USAA in our PAMI paper, we are using the following settings: We use three splits for ZSL, and the zero-shot (testing classes) are [1,2,4,7];[1,6,7,8],[2,4,5,6]; Since we provide instance-level for each video, the binary class-level prototype should be
>
> $mean(attribute(video == classname, :)) > Threshold;$
>
> i.e. the mean of the video attributes belong to the same class; And the threshold=0.5. More related work:
>
> On AwA dataset, please refer to the evaluation methods in [16] or [23]. Either one is OK. It also provides some samples codes from their websites AwA1[9]. Note that you should write clearly in your report which evaluation metrics are used in your paper.

### 2.3 Large-scale video classification

Videos carry very rich and complex semantics, and are intrinsically multimodal. Techniques for recognizing high-level semantics in diverse unconstrained videos can be deployed in many applications, such as Internet video search. However, it is well-known that semantic recognition or categorization of videos is an extremely challenging task due to the semantic gap between computable low-level video features and the complex high-level semantics. While significant progress has been achieved in recent years, most state-of-the-art solutions rely on a large set of features with simple fusion strategies to model the high-level semantics. For instance, two popular ways of combining multiple video features are early fusion and late fusion. Early fusion concatenates all the feature vectors into a long representation for classifier training and testing, while late fusion trains a classifier using each feature separately and combines the outputs of all the classifiers. Both methods do not have the capability of explicitly modeling the correlations among the features, which can be exploited to achieve a better representation. In addition, the existing categorization methods often neglected the relationships of different semantic classes, which can be exploited to boost the categorization performance. Although there exist several works investigating multi-feature fusion or exploiting the inter-class relationships, they mostly address the two problems separately.

Our group has one big dataset for video understanding. The dataset can be downloaded from FCVID[10]. Note that low-level features have been extracted and provided in the link.

The whole dataset is well described in FCVID[11]. For this dataset, please read [12].

[8] `http://yanweifu.github.io/USAA/download/`
[9] `https://cvml.ist.ac.at/AwA/`
[10] `http://bigvid.fudan.edu.cn/data/fcvid/`
[11] `http://bigvid.fudan.edu.cn/FCVID/`

#### 2.3.1 Minimum requirements

For this task, one can only finish the task of supervised learning for all 239 classes. The minimum requirements include:

1. randomly sampling the training instances of each video classes from few training instances to large number of training instances;
2. comparing different types of features; and discuss the differences and complementarity of each type of features.
3. exploring the relationships of all classes as Sec.4.2.2 in [12].
4. You need to compare different fusion strategies, including early fusing, and late fusion.
5. Note that we do care about the classification performance on this task.

### 2.4 Identification and Synthesizing of Raphael's paintings from the forgeries

The following data, provided by Prof. Yang WANG from HKUST[12]. Since this link is in google drive, you can download the file from our course webpage[13]. The data contains a 28 digital paintings of Raphael or forgeries. Note that there are both jpeg and tiff files, so be careful with the bit depth in digitization. The The following file[14] contains the labels of such paintings, which are

Now, an updated list of what is what:

1. Maybe Raphael - Disputed
2. Raphael
3. Raphael
4. Raphael
5. Raphael
6. Raphael
7. Maybe Raphael - Disputed
8. Raphael
9. Raphael
10. Maybe Raphael - Disputed
11. Not Raphael
12. Not Raphael
13. Not Raphael

[12] https://drive.google.com/folderview?id=0B-yDtwSjhaSCZ2FqN3AxQ3NJNTA&usp=sharing
[13] http://www.sdspeople.fudan.edu.cn/fuyanwei/course/projects/final_project/Raphael_Project_img.zip
[14] https://docs.google.com/document/d/1tMaaSIrYwNFZZ2cEJdx1DfFscIfERd5Dp2U7K1ekjTI/edit

14. Not Raphael
15. Not Raphael
16. Not Raphael
17. Not Raphael
18. Not Raphael
19. Not Raphael
20. My Drawing (Raphael?)
21. Raphael
22. Raphael
23. Maybe Raphael - Disputed
24. Raphael
25. Maybe Raphael - Disputed
26. Maybe Raphael - Disputed
27. Raphael
28. Raphael

#### 2.4.1 Questions

1. Can you exploit the known Raphael vs. Not Raphael data to predict the identity of those 6 disputed paintings (maybe Raphael)? The following student poster report seems a good exploration[15]. The paper [16] by Haixia Liu, Raymond Chan, and me studies Van Gogh's paintings which might be a reference for you.

2. We need to synthesize the painting of Raphael. That is, given one photo, we need to generate/synthesize a new photo which makes it look like from Raphael. There are lots of recent works on this topic. You can try it by either (1) Generative Adversarial Networks [11], or Convolutional Neural Networks [9], or Attribute transfer [18]. The testing images are downloaded from [17]. You can use the images in [18] as well as other available images online to train your model. Your report should show the effects of synthesized testing images.

#### 2.4.2 Minimum requirements

The minimum requirements include:

1. classification tasks in Question-1;
2. synthesizing tasks in Question-2.

[15] http://www.sdspeople.fudan.edu.cn/fuyanwei/course/projects/final_project/artistic_poster.pdf
[16] http://dx.doi.org/10.1016/j.acha.2015.11.005
[17] http://www.sdspeople.fudan.edu.cn/fuyanwei/course/projects/final_project/test_images.zip
[18] http://www.sdspeople.fudan.edu.cn/fuyanwei/course/projects/final_project/Raphael_Project_img.zip

### 2.5 Other projects.

You can also try other projects. However, please let TA and me know first in order to get the approval. Note that if the project is too easy, it would affect your scores of final projects.

## References


[1] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Label-embedding for attribute-based classification. In *CVPR*, 2013.

[2] Zeynep Akata, Scott Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. Evaluation of output embeddings for fine-grained image classification. In *CVPR*, 2015.

[3] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. Describing objects by their attributes. In *CVPR*, 2009.

[4] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE TPAMI*, 2006.

[5] Andrea Frome, Greg S. Corrado, Jon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. DeViSE: A deep visual-semantic embedding model. In *NIPS*, 2013.

[6] Yanwei Fu, Timothy Hospedales, Tao Xiang, and Shaogang Gong. Attribute learning for understanding unstructured social activity. In *ECCV*, 2012.

[7] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Zhengyong Fu, and Shaogang Gong. Transductive multi-view embedding for zero-shot recognition and annotation. In *ECCV*, 2014.

[8] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, and Shaogang Gong. Learning multi-modal latent attributes. *IEEE TPAMI*, 2013.

[9] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*. 2017.

[10] T. Hertz, A.B. Hillel, and D. Weinshall. Learning a kernel function for classification with small training samples. In *ICML*, 2016.

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial nets. In *CVPR*. 2017.

[12] Yu-Gang Jiang, Zuxuan Wu, Jun Wang, Xiangyang Xue, and Shih-Fu Chang. Exploiting feature and class relationships in video categorization with regularized deep neural networks. In *IEEE TPAMI*, 2017.

[13] Ricardo JVilalta and Youssef. Drissi. A perspective view and survey of meta-learning. *Artificial intelligence review*, 2002.

[14] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML – Deep Learning Workshok*, 2015.

[15] Brenden M. Lake and Ruslan Salakhutdinov. One-shot learning by inverting a compositional causal process. In *NIPS*, 2013.

[16] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009.

[17] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE TPAMI*, 2013.

[18] Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. Visual attribute transfer through deep image analogy. In *ACM Siggraph*. 2017.

[19] Marcus Rohrbach, Sandra Ebert, and Bernt Schiele. Transfer learning in a transductive setting. In *NIPS*, 2013.

[20] S. Thrun. *Learning To Learn: Introduction*. Kluwer Academic Publishers, 1996.

[21] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In *NIPS*, 2016.

[22] Jason Weston, Samy Bengio, and Nicolas Usunier. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, 2011.

[23] Yongqin Xian, Bernt Schiele, and Zeynep Akata. Zero-shot learning - the good, the bad and the ugly. In *CVPR*, 2017.