Machine Learning Basics

A self-study metarials for PRML [1]

Jisung Lim[1]

[1] B.S. Candidate of Industrial Engineering
Yonsei University, South Korea.

28th January, 2017

Summary

YONSEI UNIVERSITY

Optical Character Recognition

OCR Problem and Some Approaches

FIGURE 1: The MNIST database.

- **Input** : 28×28 pixel image, represented by a vector $\boldsymbol{x}$ comprising 784 real numbers.
- **Goal** : To build a machine that will take such a vector $\boldsymbol{x}$ as input and that will produce the identity of the digit $0, \ldots, 9$ as the output.

Consider the example of recognizing handwritten digits, illustrated in Figure 1. This is the nontrivial problem due to the wide variability of handwriting. And we can tackle this problem using following approches :

1. Handcrafted rules or heuristics
2. Machine learning methods

In practice, the former leads to a proliferation of rules and invariably gives poor results. For better results, the latter can be considered an alternative.

YONSEI UNIVERSITY

Machine Learning Approach

Machine Learning Approach A machine learning approach can be divided into three major stages : **training**, **testing**, and **predicting**.

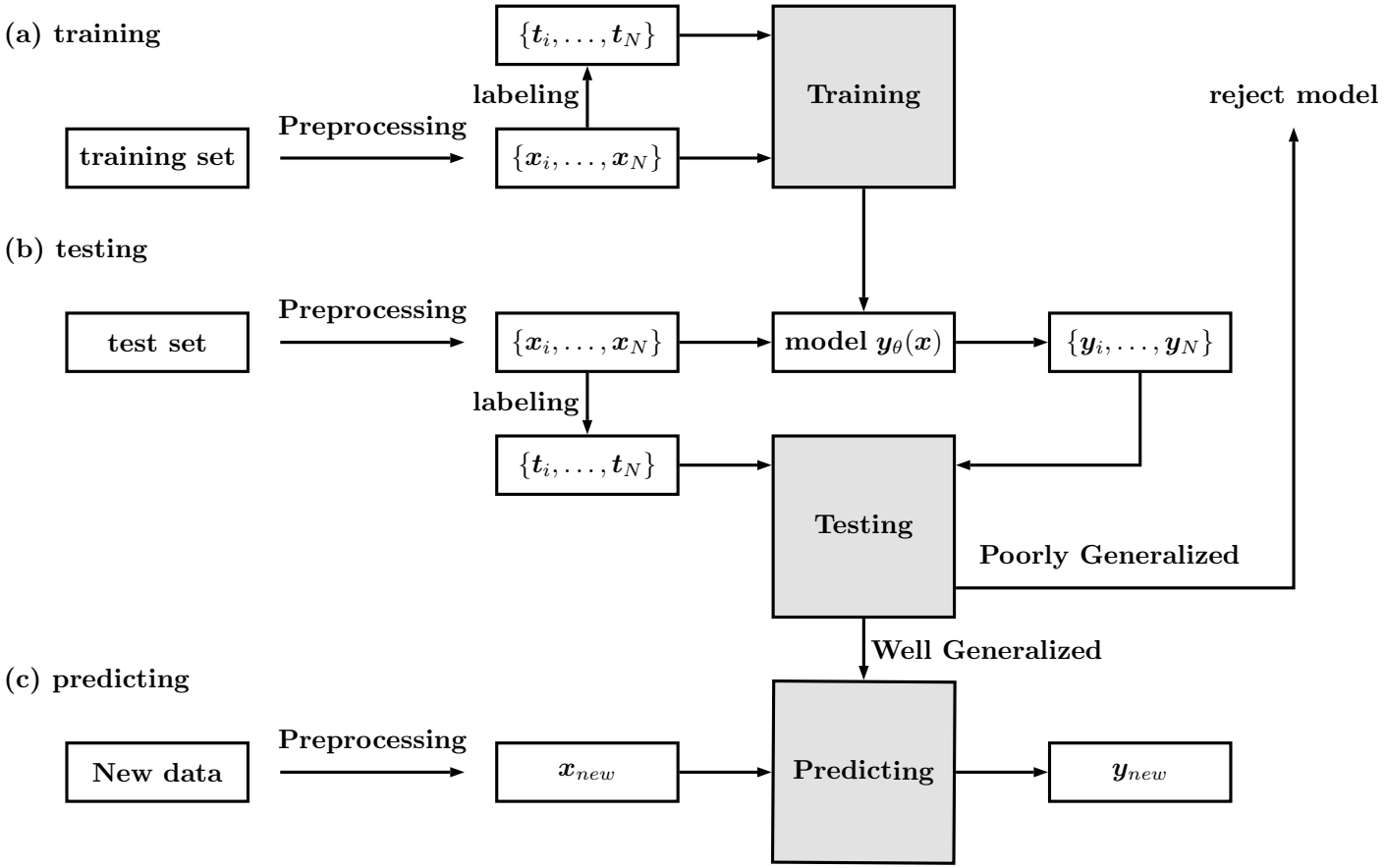

FIGURE 2: The description of full machine learning approach.

ML Approach — Input

First of all, we need to define inputs more precisely. In OCR, inputs are handwriting images of a digit. In practical applications, the original input images may have different sizes, ratios, angles, or even colors. Hence, the images are typically scaled, rotated, translated, and grey-scaled so that each digit is contained within a box of a fixed-size, says 28×28, and is toned with greyscale color. And then, the preprocessed input images can be represented as a vector $\boldsymbol{x} = (x_1, \ldots, x_{784}) \in \mathbb{R}^n$ where the greyscale color of ith pixel is a x_i. Finally, you should label the true number $\boldsymbol{t}$ for each input image $\boldsymbol{x}$.

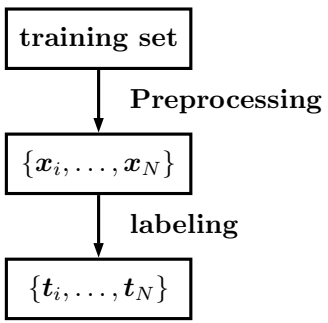

FIGURE 3

Preprocess : To transform the original input variables into some new space of variables so that the problem becomes easier to solve.

Example x_i : An example is a collection of features. Typically, an **example** will be represented as a vector $\boldsymbol{x}_i \in \mathbb{R}^n$ where each entry x_i is a **feature**.

Label t_i : A **label** represents the identity of the corresponding digit. Typically, the labels are hand-labelled by inspecting each image individually. Note that there is one such **label** $\boldsymbol{t}_i$ for each **example** $\boldsymbol{x}_i$.

YONSEI UNIVERSITY

ML Approach — Train and Test

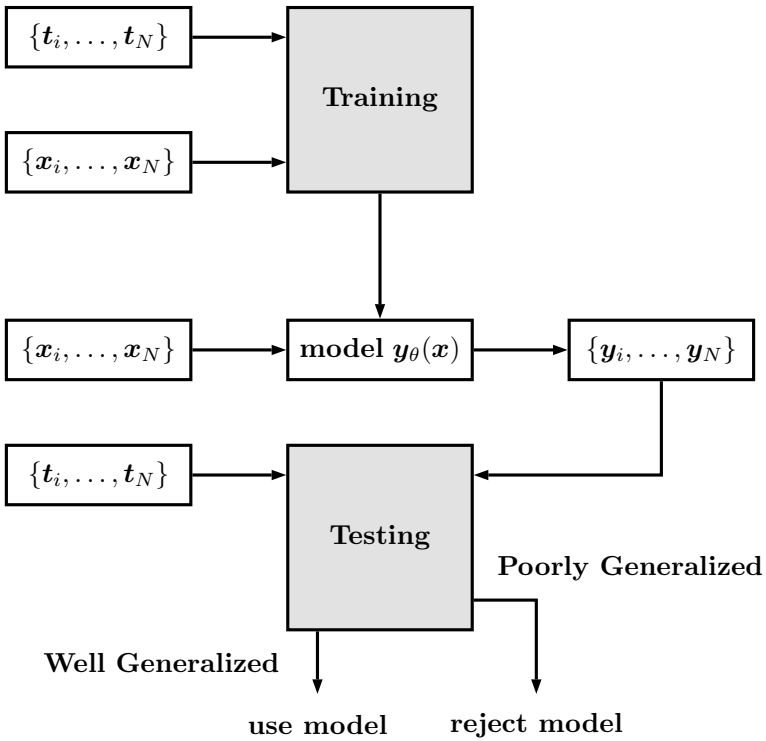

FIGURE 4: The description of training and test.

Training

To tune the parameters of an adaptive model, it uses a large set of N **examples** $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ with its corresponding **labels** $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N\}$. The model can be expressed as a function $\boldsymbol{y}(\boldsymbol{x})$ which takes a new digit image $\boldsymbol{x}_{new}$ as input and that generates an output vector $\boldsymbol{y}_{new}$.

Testing

Once the model is trained it can then determine the identity of new digit images, which are said to comprise a test set. The ability to categorize correctly new examples that differ from those used for training is known as generalization. In practical applications, the variability of the input vectors will be such that the training data can comprise only a tiny fraction of all possible input vectors, and so generalization is a central goal in patter recognition.

YONSEI UNIVERSITY

ML Approach — Prediction

Now the model can works for new, unlabeled inputs. In OCR, prediction stage can be described as below.

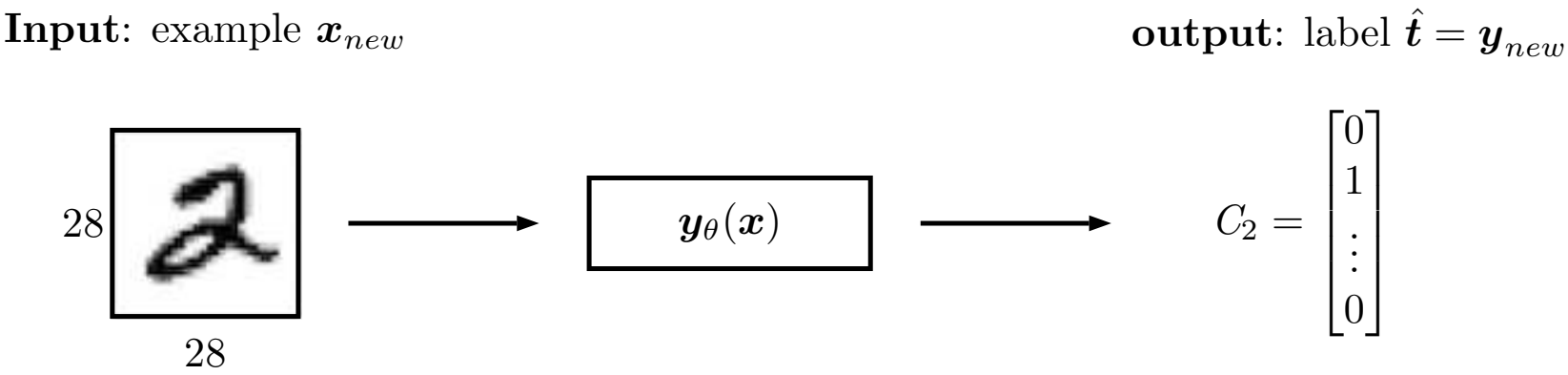

FIGURE 5: The description of prediction in OCR.

ML Approach — SL, UL, and RL

SL Supervised Learning :
Corresponding target vectors are known.
Ex. Regression, Classification.

UL Unsupervised Learning :
The training data consists of a set of input vector $\boldsymbol{x}$ without any corresponding target values.
Ex. Clustering, Density estimation.

RL Reinforcement Learning :
Finding **suitable actions** to take in a given situation in order to maximize a reward.
It discovers the optimal by a process of **trial and error**.
A general feature of reinforcement learning is the trade-off between **exploitation**, in whcih the system **tries out new kinds of actions** to see how effective they are, and **exploitation**, in which the system **makes use of actions** that are known to yield a high reward.

Precise Definition of ML

Definition of machine learning

A computer program is said to ***learn*** from **experience** ***E*** with respect to some class of **tasks** ***T*** and **performance measure** ***P***, if its performance at tasks in ***T***, as measured by ***P***, improves with experience ***E***.

Polynomial Curve Fitting

Given Situation

- N observations $\mathbf{X} \equiv (x_1, \ldots, x_N)^T$, together with corresponding labels $\mathbf{t} \equiv (t_1, \ldots, t_N)^T$
- (x_i, t_i) possess underlying regularity, which we wish to learn.
- Individual t_i observations are corrupted by random noise, which might arise from intrinsical stochasticity due to there being source of variability.

Our Goal and Intrinsic Difficulty

- **Goal**
  exploit this given training set in order to make prediction of the value $\hat{t}$ of the target variable for some new value $\hat{x}$ of the input variable.
- **Intrinsic Difficulty**
  This is intrinsically a difficult problem not only as we have to generalize from a **finite data** set but also, for a given $\hat{x}$ there is **uncertainty** as to the appropriate value for $\hat{t}$.

1. minimizing an error function.

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(\boldsymbol{x}_n, \boldsymbol{w}) - \boldsymbol{t}_n\}$$

$$where\ y(x, \boldsymbol{w}) = w_0 + w_1 x + w_2 x^2 +$$

2. use regularization for preventing overfitting.

$$\tilde{E}(\boldsymbol{w}) = E(\boldsymbol{w}) + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{w}$$

YONSEI UNIVERSITY

Chapter Objectives

- **Probability Theory**
  Provides a consistent framework for expressing such uncertainity in a precise and quantitative manner.
- **Decision Theory**
  allows us to exploit this probabilistic representation to make optimal prediction according to appropriate criteria.
- **Information Theory**


YONSEI UNIVERSITY

 Christopher M. BISHOP. **Pattern Recognition and Machine Learning**. Springer, 2006.