**Chapter 5**

**Machine Learning Approach**

**5.1 Artificial Intelligence**

Research in Artificial Intelligence (AI) is very general and spawns across numerous different area such as mathematics, computer science, philosophy, economics and even ethics. This field is very wide and can be attempted from many different view points. Therefore, this explanation will not be very exhaustive. For comprehensive and more complex joint into subject refer to [28].

One of many possible definitions of AI can be brief as a search to develop an artificial intelligent agent. In other words, it is an effort to create intelligent machines that are either intelligent or can be perceived as intelligent ones. One of the most important skills of intelligent agents is the sense of vision. Sense of vision is usually required to positive degree and not always it is necessary that it rivals the abilities of human visual apparatus.

First tries to resolve the vision problems were tackled from so called bottom-up approach in which the system was instructed with hard-coded set of protocol describing the vision. It was expected that as the understanding of a instruments allowing humans to abstract information from visual scene, the hard-coded systems can be fed this understanding and thus more skilled systems can be created. Problem with this method was that it highly underestimated the difficulty of reinforcement of these protocols. Therefore, it mainly failed to devise such a system.

This main idea the investigators to postulating that in order to resolve the problem of deploying vision capabilities for artificial intelligent system, it is compulsory to introduce a procedure that would allow AI systems to extract patterns from provided data. It is an overview of systems that can learn. Process that enables systems to learn is usually called machine learning.

Machine learning is again relatively extensive term that can be used in multiple different frameworks. In this work, it is meant to be understood as technique that is used to create mathematical representations used for image detection. There are numerous types of machine learning models that are useful for different tasks. The task that is conversed in this work and is also arguably most commonly attempted is called classification, which is the task to classify occurrence of input into correct discreet and mainly predetermined class. One more common type of machine learning task is called regression, which is based on the input data trying to estimation unknown continues valued quantity.

Rest of this document will be exclusively dealing with learning tasks of classification type.

**5.2 Image Processing**

Computer vision is important topic in image processing. Investigation in the subject was approaching the problem from bottom-up perception for many years. This was the before-mentioned effort to validate directions guiding the vision of living organisms. This method was actually very successful in certain environments.

General explanation of computer vision in image processing can be brief in following steps:

• Image capture - Image is captured using camera or similar device and digitized.

• Pre-processing - Captured image is modified to highlight important features such as noise reduction, contrast normalization etc.

• Segmentation detection - Selection of region of interest like edges, similar surfaces.

• Description – Feature extraction of radiometric, photometric descriptors and so on.

• Classification - Some means to categorize the object.



Figure 5.1: Diagram of image processing pipeline. [20]

Specific steps are shown in figure 5.1. Even though machine learning been as field of study since the second half of the 20th century, there was no wider implementation of its techniques in image processing for very long time. It was first announced in the classification step of the processing pipeline. In other words, complex difficulties were simplified by reducing the visual information contained within the image into handful of simple features that were nourished into machine learning model. This method consequently favors very simple models such as K-nearest Neighbors (KNN) or Support Vector Machines (SVM).

This carries the problem that these applications are not very useful. Each application is usually only capable of resolving very narrow problem and any deviation from ideal circumstances can mean failure. Application can have complications with varying contrast, brightness, scaling, rotation etc.

Second problem is often the fact that because image must be pre-processed numerous times before it is fed into machine learning model it requires additional time and computation resources. This is less of a problem with existing hardware innovations, but it is still not unimportant factor and it can have negative effect on the cost of the result. This is where machine learning in general indicates noteworthy advantage.

Conventional approach to computer vision can find success in applications that are deployed in very limited environments with rigid constraints. In controlled location, it is usually very simple to define the problem in formal guidelines. Even though it can be still availabel in certain places, it starts to be forced out by application of machine learning simply because the barrier of entry for extensive adoption reductions every day.

**5.3 Machine Learning**

As before described, Machine Learning is a process that is used to create models that can abstract information from data to resolve given problem and consequently repeatedly improves their performance.

Interesting viewpoint that can be used is to view machine learning as form of information compression. Where machine learning model is trying to abstract information from input data in such a way that the amount of a data used to save is summary while the information contained within is preserved.

**5.3.1 Machine Learning Approach**

There are mainly two different types of machine learning methods:

• Unsupervised Learning

• Supervised Learning

Both are mainly used for different types of machine learning tasks.

**Unsupervised learning**

In unsupervised learning method, the model is training by detecting new data and take out patterns in the date without being instructed on what they are. Contrasting to supervised learning defined below, the benefit of this method is that the model is able to learn from data without supervision. This means that there is no need for input data to be explained, therefore it takes a smaller amount time and resources to deploy these models in practice.

The biggest difficulty of supervised learning method in real world applications is to obtain appropriate data. Appropriate data in this context means, data that were someway classified into different categories, which can be very boring and slow process. In some situation, the task itself avoids the usage of labeled data. Labeled data are impossible to find or don’t exist at all.

Mainstream of unsupervised learning procedures belong to group called clustering algorithms. These algorithms are centered on the idea to analyze ordered clustering of data in input space to determine their relationship. This is achieved by the belief that data point clustering in input space are likely to exhibit like properties.

Illustrations of unsupervised learning models are:

• K-means -clustering model [17];

• Self Organizing Maps (SOMs) [22];

• Principal Component Analysis (PCA) - dimensionality reduction [17].

Image classification usually does not depend on the use of unsupervised learning methods; therefore the following writing describes only supervised learning methods.

**Supervised learning**

Supervised learning method is more commonly used. This method needs training data with specific format. Each instance must have assigned label. These labels make available supervision for the learning algorithm. Training process of supervised learning is constructed on the following principle. First, the training data are fed into the model to produce estimate of output. This estimate is compared to the assigned label of the training data in order to evaluation model error. Based on this error the learning algorithm alters model’s parameters in order to reduce it.

**5.3.2 Structure of Machine Learning Algorithm**

Although machine learning algorithms are various and are using different methods its structure can be generalized. Structure of nearly all machine learning algorithms can be defined as composition of the following components:

• Dataset description

• Model

• Cost function

• Optimization technique

Almost all supervised learning algorithms use the same Dataset description. The other three components can vary intensely. This level of analysis is suitable for building of intuition for Neural Networks (NNs) and description of its individual components.

**Dataset description**

Supervised learning requires datasets with detailed properties. Each dataset holds set of 𝑛 instances which contains of a pair of input vector 𝑥𝑖 and output scalar 𝑦𝑖. Input vector

𝑥 𝑇 𝑖 = [𝑥1, 𝑥2, . . . , 𝑥𝑝], (5.1)

where 𝑖 is index of instance, 𝑝, is length of input vector.

Specific components of input vector have to be of unified type. In case of image as input data it is value for individual pixels (e.g. 0-255). In other cases, they could be real values. Almost commonly in machine learning it stands that input should be normalized. This belief holds in images automatically since each pixel must have its vales in fixed range. It is very significant in other types of machine learning tasks, where this is not guaranteed.

Output scalar 𝑦𝑖 characterises class of given instance. Type of this output value thus must obtain only certain values. To put it differently, it must be a set of cardinalities equal to number of all possible classes.

**Model**

Model is prediction tackle that takes input 𝑥𝑖 to predict value of its output 𝑦𝑖. To each model has parameters represented by vector 𝜃, which are used to during the training process. The modest example of model type is linear model, also called linear regression. Parameters 𝜃 of this model are

𝜃 𝑇 = [𝜃1, 𝜃2, . . . , 𝜃𝑝], (5.2)

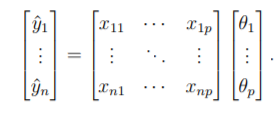
where 𝑝 is number of parameters equal to size of input vector 𝑥𝑖 . Prediction 𝑦^𝑖 of the model on instance 𝑖 is computed as

𝑦^𝑖 = ∑︁ 𝑝 𝑗=1 𝑥𝑖𝑗𝜃𝑗. (5.3)

Therefore estimates of the model on the entire dataset in matrix notation is

𝑦^ = 𝑋𝜃. (5.4)

Estimates in expanded notation are equal.



**Cost Function**

To achieve the learning skill of the machine learning algorithm, it is necessary to approximation the error of its predictions. This is assessed with so called cost function and called loss function.

This function must have certain properties. Ability of the machine learning algorithm to learn rests on the approximation of its improvement with change of its parameters. Therefore, cost function must be at least partially differentiable. In case of linear regression, it is most common to usage sum of square error. The main aim being that derivative of this function for linear model has only one global minimum.

Cost function is defined as

𝐽(𝜃) = ∑︁𝑛 𝑖=1 (𝑦𝑖 − 𝑦^𝑖) 2 = ∑︁𝑛 𝑖=0 (︁ 𝑦𝑖 − 𝑥𝑖 𝑇 𝜃 )︁ 5 . 6

For the optimization determinations it is usually useful to express the cost function in matrix notation

𝐽(𝜃) = (𝑦 − 𝑋𝜃) 𝑇 (𝑦 − 𝑋𝜃). (5.7)

**Optimization technique**

The last part of learning algorithm is the optimization technique. It consists of update of model’s parameters 𝜃 in order to progress its prediction. In other words, to find 𝜃 such that the value of cost function 𝐽(𝜃) for given dataset is as small as possible.

To examine the change of cost function on given dataset it is necessary to compute the derivative of 𝐽(𝜃) with respect to 𝜃

𝜕𝐽(𝜃) 𝜕𝜃 = 𝜕 𝜕𝜃 [︁ (𝑦 − 𝑋𝜃) 𝑇 (𝑦 − 𝑋𝜃) ]︁

= 𝜕 𝜕𝜃 [︁ 𝑦 𝑇 𝑦 + 𝜃 𝑇𝑋𝑇𝑋𝜃 − 2𝑦 𝑇𝑋𝜃]︁

= 2𝑋𝑇𝑋𝜃 − 2𝑋𝑇 𝑦. (5.8)

For linear model is possible to find optimal result which is global minimum of the cost function. The optimal result

𝜃 = (︁ 𝑋𝑇𝑋 )︁−1 𝑋𝑇 𝑦, (5.9)

is found by comparing the partial derivative of 𝐽(𝜃) to 0. Only condition is that 𝑋𝑇𝑋 must be non-singular.

Unfortunately, only very simple problems can be like using model as simple as linear regression. More complex model usually means more complex cost function. Optimization process of more complex cost functions cannot be definite to find global minimum. In this case, the optimization technique must be of iterative character. To put it in a dissimilar way, algorithm has to method the minimum of iterations. Many of the iterative approaches belong to the group called gradient based optimization.

**5.3.3 Model Complexity**

In the first calculation it could be said that the task of supervised machine learning is to model relationship between the input output data most correctly. The problem with this definition is that in the real-world application there is never enough data to capture true relationship between the two. Therefore, the task of machine learning is the attempt to suppose true relationship by detecting incomplete picture.

Hence the most significant property of machine learning model is its generalization capability. That is capability to produce meaningful outcomes from data that were not previously detected.

Generalization ability is reliant on complexity of the model and its relationship to complexity of underling problem. When model does not capture complexity of the problem appropriately it is defined as under fitting. In case the complexity of model surpasses the complexity of underling problem then this phenomenon is called over fitting.

In both extremes the generalization ability suffers. In the earlier case the model is unable to capture true intricacies of the problem and therefore is unable to predict wanted output reliably. In the last case it tries to capture even the subtlest data perturbation that might be in fact an outcome of stochastic nature of the problem and not the real underlying relationship. This can also cause the fact that input data is lost some variable necessary to capture the true relationship. This fact is unescapable, and it thus must be careful when designing machine learning model. Representation of this phenomena in case of two variable inputs is on Figure 5.2.



Figure 5.2: Figure shows different levels of generalization of model [1]

Typically, the machine learning model is trained on as much input data as possible in order to reach the best possible performance. At the same time its error rate must be verified on independent input data to checked whether the generalization ability is not deteriorating. This is typically accomplished by splitting available input data into training and testing set. Frequently in 4:1 fraction for training to test data. Model is trained with training data only and the presentation of the model is tested on the test data. Connection between test and train error can be found on Fig. 5.3. Even though the true generalization error can never be truly detected, its estimate by test error rate is enough for majority of machine learning tasks.

**Regularization**

Regularization is any alteration that is made to the learning algorithm that is intended to decrease its generalization error but not its training error [16]. As it has already been stated, the most significant aspect of machine learning is striking the stability between over and under fitting of the model. To support with this problem concept of regularization was devised. It is a method that helps to penalize the model for its complexity. Basic idea consists of adding a term in the cost function that increases with model complexity. When this is applied to cost function from equation 5.7

𝐽(𝜃) = (𝑦 − 𝑋𝜃) 𝑇 (𝑦 − 𝑋𝜃) + 𝜆𝜃 𝑇 𝜃, (5.10)

where 𝜆 is a parameter that controls the strong point of the preference [16].



Figure 5.3: Relationship between the model complexity and its ultimate accuracy is the relationship between training and testing error [7].