**2.1 Artificial Intelligence**

Field of Artificial Intelligence (AI) is very general and spawns across several different disciplines (e.g. mathematics, computer science, philosophy, economics and even ethics). This fact suggests that this field is very broad and can be tackled from many different view points. Therefore, this description will not be very exhaustive. For broader and more complex dive into subject refer to [28].

One of many possible definitions of AI can be summarized as a pursuit to develop an artificial intelligent agent. In other words, it is an attempt to create intelligent machines that are either thinking or can be perceived as thinking ones. One of the most important abilities of intelligent agents is the sense of vision1 . Sense of vision is usually required to certain degree and not always it is necessary that it rivals the capabilities of human visual apparatus.

First attempts to solve the vision problems were tackled from so called bottom-up approach in which the system was instructed with hard-coded set of rules describing the vision. It was expected that as the understanding of a mechanisms allowing humans to extract information from visual scene, the hard-coded systems can be fed this understanding and thus more capable systems can be created. Problem with this approach was that it highly underestimated the difficulty of formalization of these rules. As a consequence, it predominantly failed to devise such a system.

This insight leads the researchers to postulating that in order to solve the problem of deploying vision capabilities for artificial intelligent system, it is necessary to introduce a process that would allow AI systems to extract patterns from provided data. That is to say, it is an introduction of systems that are able to learn. Process that enables systems to learn is generally called machine learning.

Machine learning is again quite extensive term that can be used in multiple different contexts. In this work, it is meant to be understood as technique that is used to create mathematical models used for image recognition. There are several types of machine learning models that are useful for different tasks. The task that is discussed in this work and is also arguably most commonly tackled is called classification, which is the task to classify instance of input into correct discreet and mainly finite class. Another common type of machine learning task is called regression, which is based on the input data trying to estimate unknown continues valued quantity.

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

**2.2 Image Processing**

Computer vision is important subject in image processing. Research in the subject was approaching the problem from bottom-up perspective for many years. This was the before-mentioned effort to formalize rules guiding the vision of living organisms. This approach was actually very successful in certain circumstances.

General description of computer vision in image processing can be summarized in following steps:

• Image capture - Image is captured (by camera or similar device) and digitized.

• Pre-processing - Digitized image is modified to emphasize important features (e.g. noise reduction, contrast normalization etc.).

• Segmentation - Selection of interesting features (edges, similar surfaces).

• Description - Extraction of radiometric, photometric descriptors and so on.

• Classification - Some means to classify the object.



Figure 5.1: Block diagram of image processing pipeline. [20]

Individual steps are shown in figure 2.1. Even though machine learning existed as field of study since the second half of the 20th century, there was no wider adoption of its techniques in image processing for very long time. It was first introduced in the classification step of the processing pipeline. In other words, complex problems were simplified by reducing the visual information contained within the image into handful of simple features that were fed into machine learning model. This approach consequently favors very simple models such as K-nearest Neighbors (KNN) or Support Vector Machines (SVM).

This brings the problem that these applications are not very versatile. Each application is usually only capable of solving very narrow problem and any deviation from ideal circumstances can mean failure. Application can have problems with varying contrast, illumination, scaling, rotation etc.

Second problem is often the fact that because image has to be pre-processed several times before it is fed into machine learning model it requires extra time and computation resources. This is less of a problem with current hardware innovations, but it is still not negligible factor and it can have negative effect on the cost of the solution. This is where machine learning in general indicates significant advantage.

Classical approach to computer vision can find success in applications that are deployed in very restricted environments with rigid constraints2 . In controlled environment, it is usually very simple to describe the problem in formal rules. Even though it can be still found in certain places, it starts to be forced out by application of machine learning simply because the barrier of entry for wider adoption decreases every day.

**2.3 Machine Learning**

As previously described, Machine Learning is a process that is used to create models that are able to extract information from data to solve given problem and consequently automatically improves their performance.

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

**2.3.1 Machine Learning Approach**

There are typically two different types of machine learning approaches:

• Unsupervised Learning

• Supervised Learning

Both are typically used for different kinds of machine learning tasks.

**Unsupervised learning**

In this learning approach, the model is training by observing new data and extracting patterns in the date without being instructed on what they are. Opposed to supervised learning described below, the advantage of this approach is that the model is able to learn from data without supervision (as the name suggests). This means that there is no need for input data to be annotated, therefore it takes much less time and resources to deploy these models in practice.

The biggest hurdle of supervised learning approach in real world applications is to obtain appropriate data. Appropriate data in this context means, data that were somehow classified into different categories, which can be very tedious and slow process. In some cases, the task itself prevents the usage of labeled data (i.e. labeled data are impossible to obtain or don’t exist at all).

Majority of unsupervised learning algorithms belong to group called clustering algorithms. These algorithms are centered on the idea to analyze geometric clustering of data in input space to determine their affiliation. This is achieved by the presupposition that data point clustering in input space are likely to exhibit similar properties.

Examples of unsupervised learning models are:

• K-means - clustering model [17, p. 460–462];

• Self Organizing Maps (SOMs) - instance based [22];

• Principal Component Analysis (PCA) - dimensionality reduction [17, p. 534– 544].

Image classification usually does not rely heavily on the use of unsupervised learning methods, therefore the following text describes only supervised learning methods.

**Supervised learning**

Supervised learning approach is more commonly used. This approach requires training data with specific format. Each instance has to have assigned label. These labels provide supervision for the learning algorithm. Training process of supervised learning is based on the following principle. Firstly, the training data are fed into the model to produce prediction of output. This prediction is compared to the assigned label of the training data in order to estimate model error. Based on this error the learning algorithm adjusts model’s parameters in order to reduce it.

**2.3.2 Structure of Machine Learning Algorithm**

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

• Dataset specification

• Model • Cost function

• Optimization procedure

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

**Dataset specification**

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

𝑥 𝑇 𝑖 = [𝑥1, 𝑥2, . . . , 𝑥𝑝], (2.1)

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

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

Output scalar 𝑦𝑖 represents class of given instance. Type of this output value thus has to acquire only certain values. To put it differently, it has to be a set of cardinality equal to number of all possible classes.

**Model**

Model is prediction apparatus that takes input 𝑥𝑖 to predict value of it’s output 𝑦𝑖. Each model has parameters represented by vector 𝜃, which are adjusted during the training process. The simplest example of model type is linear model, also called linear regression. Parameters 𝜃 of this model are

𝜃 𝑇 = [𝜃1, 𝜃2, . . . , 𝜃𝑝], (2.2)

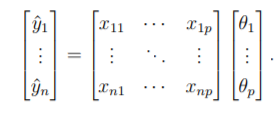
where 𝑝 is number of parameters equal to size of input vector 𝑥𝑖 . Prediction 𝑦^𝑖 of the model on instance 𝑖 is computed as

𝑦^𝑖 = ∑︁ 𝑝 𝑗=1 𝑥𝑖𝑗𝜃𝑗. (2.3)

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

𝑦^ = 𝑋𝜃. (2.4)

Predictions in expanded notation are equal.



**Cost Function**

To achieve the learning ability of the machine learning algorithm, it is necessary to estimate the error of its predictions. This is estimated with so called cost function (also sometimes called loss function).

This function has to have certain properties. Ability of the machine learning algorithm to learn rests on the estimation of its improvement with change of its parameters. Therefore, cost function has to be at least partially differentiable. In case of linear regression it is most common to use sum of square error. The main reason 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 )︁2 . 6

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

𝐽(𝜃) = (𝑦 − 𝑋𝜃) 𝑇 (𝑦 − 𝑋𝜃). (2.7)

**Optimization Procedure**

The last part of learning algorithm is the optimization procedure. It consist of update of model’s parameters 𝜃 in order to improve it’s 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 investigate 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. (2.8)

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

𝜃 = (︁ 𝑋𝑇𝑋 )︁−1 𝑋𝑇 𝑦, (2.9)

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

Unfortunately, only very simple problems can be approximated using model as simple as linear regression. More complex model usually means more complicated cost function. Optimization process of more complex cost functions cannot be guaranteed to find global minimum. In this case, the optimization procedure has to be of iterative character. To put it in a different way, algorithm has to approach the minimum of iterations. Many of the iterative methods belong to the group called gradient based optimization.

**2.3.3 Model Complexity**

In the first approximation it could be said that the task of supervised machine learning is to model relationship between the input output data most accurately. The problem with this definition is that in the practical application there is never enough data to capture true relationship between the two. Therefore, the task of machine learning is the attempt to infer true relationship by observing incomplete picture.

Hence the most important property of machine learning model is its generalization ability. That is ability to produce meaningful results from data that were not previously observed.

Generalization ability is dependent on complexity of the model and its relationship to complexity of underling problem. When model does not capture complexity of the problem sufficiently it is described as under fitting. In case the complexity of model exceeds the complexity of underling problem then this phenomenon is called over fitting.

In both of these extremes the generalization ability suffers. In the former case the model is unable to capture true intricacies of the problem and therefore is unable to predict desired output reliably. In the latter case it tries to capture even the subtlest data perturbation that might be in fact a result of stochastic nature of the problem and not the real underlying relationship. This can also cause the fact that input data is missing some variable necessary to capture the true relationship. This fact is unavoidable, and it thus has to be taken into account when designing machine learning model. Depiction of this phenomena in case of two variable inputs is on Figure 2.2.



Figure 2.2: 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 achieve the best possible performance. At the same time its error rate has to be verified on independent input data to check whether the generalization ability is not deteriorating. This is typically achieved by splitting available input data into training and testing set (usually in 4:1 ratio for training to test data). Model is trained with training data only and the performance of the model is tested on the test data. Relationship between test and train error can be found on Fig. 2.3. Even though the true generalization error can never be truly observed, its approximation by test error rate is sufficient for majority of machine learning tasks.

**Regularization**

Regularization is any modification that is made to the learning algorithm that is intended to reduce its generalization error but not its training error [16]. As it has already been mentioned, the most important aspect of machine learning is striking the balance between over and under fitting of the model. To help with this problem concept of regularization was devised. It is a technique that helps to penalize the model for its complexity. Basic concept consists of adding a term in the cost function that increases with model complexity. When this is applied to cost function from equation 2.7 𝐽(𝜃) = (𝑦 − 𝑋𝜃) 𝑇 (𝑦 − 𝑋𝜃) + 𝜆𝜃 𝑇 𝜃, (2.10) where 𝜆 is a parameter that controls the strength of the preference [16]. 18



Figure 2.3: Relationship between the model complexity and its ultimate accuracy is the relationship between training and testing error [7].