**Overview about Machine Learning (ISL chapter 1,10):**

Statistical learning, is the whole set of tools, helps in understanding data. These tools can be categorized to supervised, and unsupervised. Supervised statistical learning, uses given data (inputs), and predict, estimate their output, this method is broadly used. However, this project will primarily focus on supervised statistical learning, which is recognised as having variety of inputs, without having their corresponding supervised outputs.

Through unsupervised learning, by inspecting each element characteristics, we can reach the capacity of grouping similar data, within a certain dataset, to the same cluster(group). Its primary goal, is to discover significant aspects about the measurements. Principle Components Analysis, a type of unsupervised learning, is a useful tool for visualising data, or pre-processing. Moreover, Clustering, is another valuable technique, that simplifies the process of determining subgroups in dataset.

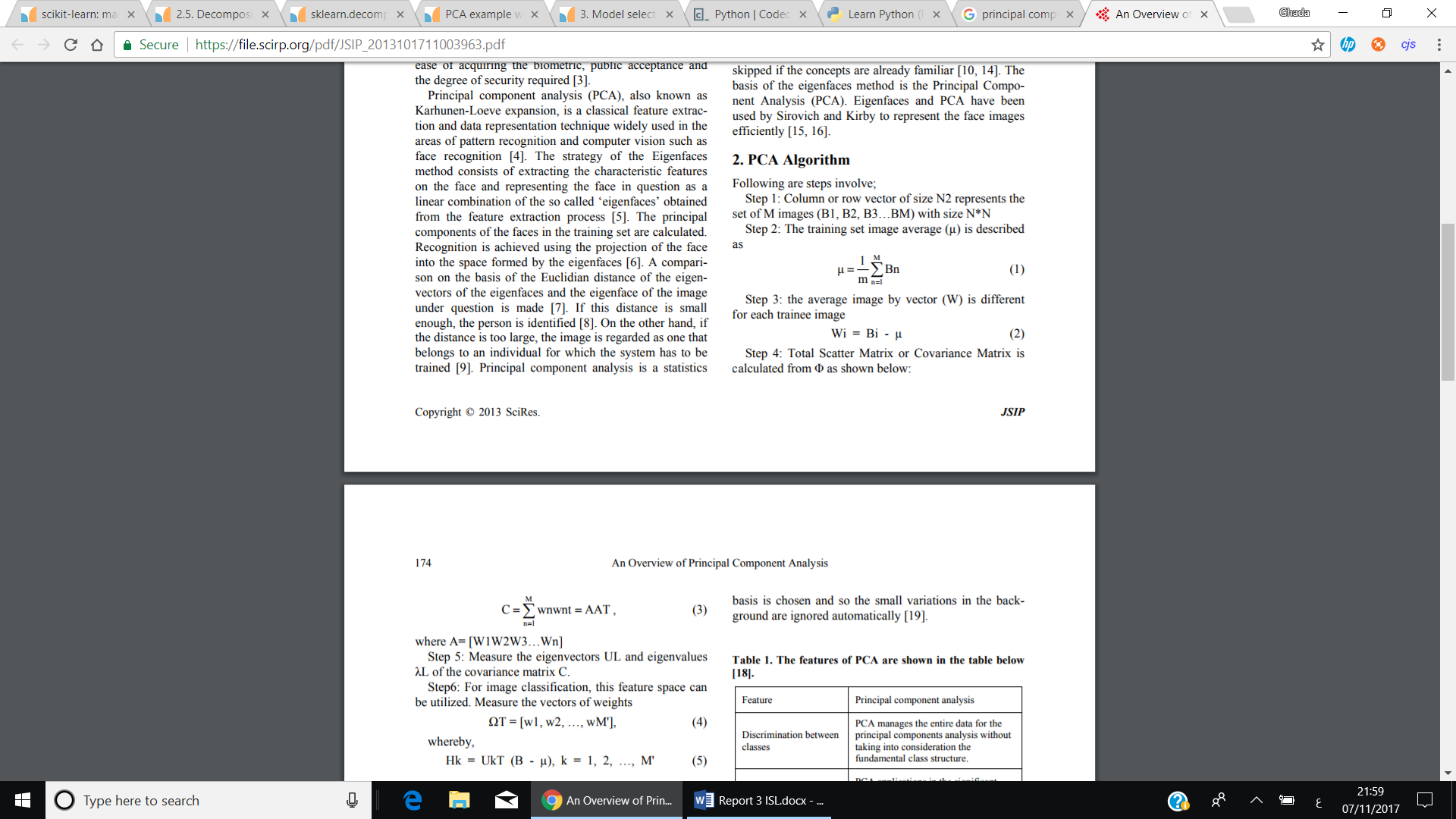
* Unsupervised Statistical Learning Tools:
  + Principal Components Analysis (PCA):

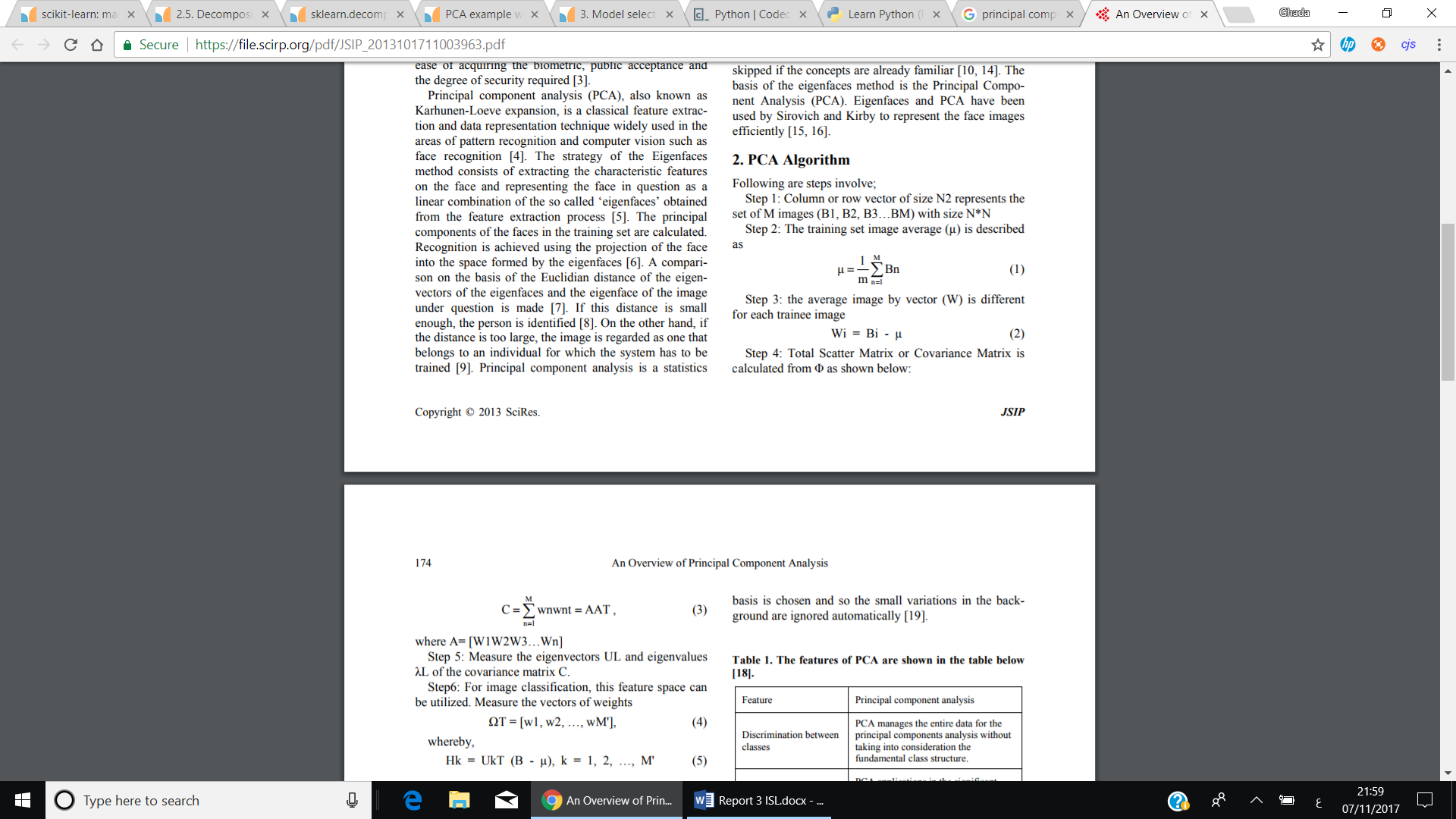
PCA Is a largely known technique, that’s used for reducing dimensionality, and decomposing a large dataset, that has a huge number of dimensions (to a lower dimensional space). The data points will be displayed, as orthogonal components, which describes the maximum amount of the variance. In general, instead of using all parameters of a dataset, using PCA will result in using, for example, 2 values at most that will eventually provide a precise representation of data.

Some of the benefits of using PCA can be summarised in few points,

* + - it clearly decreases the requirement of capacity and memory used.
    - Decrease complexity of system.
    - less sensitive to noise than other algorithms.

However, it can be argued that, evaluation of the covariance matrix result in less accuracy. Which can be viewed as a disadvantage of PCA. <https://file.scirp.org/pdf/JSIP_2013101711003963.pdf>





* 1. Scaling the variables.

Before performing PCA on observations, all data, should be scaled (using a certain method, such as standard deviation = 1).

It is important since; different variables are measured in different units, so when calculating the **variance** of each variable, it will result in first principal component loading vector to be for example, large loading, since it had the largest variance. Mostly, each variable is scaled to have unit standard deviation when the possibility of them having different units is high. Hence, scaling has a substantial impact on obtained results, and should be performed beforehand.

* 1. The minimum number of Principal Components to use is calculated via, **min(n-1, p),** where n is the number of observations and p is the number of variables. However, it is crucial that we use the very first PC’s to help in visualising data. By examining a **SCREE PLOT …**
  + Clustering:

Why Machine Learning?

Due to the significant growth of malware on devices in general, antimalware programs are trying to keep up to date with changes made. Some Antimalware are built to compare signatures to the database, to decide whether a code is considered benign or malicious, this method is now becoming less helpful with time. Using machine learning algorithms to better detect malicious code, as it has some traits that helps in deciding. “So, code that’s determined to be benign might have some traits that the software considers to be a possible indication of malware.”

Machine learning, assist devices to learn new things about data give, without the need of explicitly programming it.

For example, AI-Squared analyses large amounts of data and uses a recurrent neural network and machine learning techniques (also known as unsupervised learning) to find anomalies. Once something abnormal is found, a human analyst is then alerted to confirm whether the activity is a hacker or a genuine user. After the intent of the activity is decided, artificial intelligence takes these findings and puts them into the equation for future reference. This security process is called supervised learning.

<https://www.ccsinet.com/blog/what-to-expect-from-ai-and-cyber-security-roles-in-the-future/>

The only things that unsupervised learning methods have to work with are the observed input patterns xi, which are often assumed to be independent samples from an underlying unknown probability distribution PI [x], and some explicit or implicit a priori information as to what is important. One key notion is that input, such as the image of 1 a scene, has distal independent causes, such as objects at given locations illuminated by particular lighting. Since it is on those independent causes that we normally must act, the best representation for an input is in their terms. Two classes of method have been suggested for unsupervised learning. Density estimation techniques explicitly build statistical models (such as BAYESIAN NETWORKS) of how underlying causes could create the input. Feature extraction techniques try to extract statistical regularities (or sometimes irregularities) directly from the inputs. Unsupervised learning in general has a long and

<http://www.gatsby.ucl.ac.uk/~dayan/papers/dun99b.pdf>