Assignment #4

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari
7072310
raad00002@stud.uni-saarland.de

Dhimitrios Duka
7059153
dhdu00001@stud.uni-saarland.de

3 Problem 3 (Dimensionality Reduction)

3.1 What information does the first principal component capture in terms of the data variance and the data explaining?

The first principal component is the direction which captures the most of the variance of the data as well as at the same time, it's the closes line to the data.

3.2 Calculate the first principal component.

We can imagine the provided data as a matrix where each row represents a data point and each column represents a feature.

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{pmatrix}$$

First, we have to normalize the data. To perform normalization, first we compute the per feature mean μ_i and per feature standart deviation σ_i.

$$\begin{aligned} \mu_1 &= \frac{1+2+3}{3} = 2 \\ \mu_2 &= \frac{1+2+3}{3} = 2 \\ \sigma_1 &= \sqrt{\frac{1}{3}(1+0+1)} = 0.8164 \\ \sigma_2 &= \sqrt{\frac{1}{3}(1+0+1)} = 0.8164 \end{aligned} \tag{1}$$

Applying the normalization formula, we get the following matrix:

$$\begin{pmatrix} -1.225 & -1.225 \\ 0 & 0 \\ 1.225 & 1.225 \end{pmatrix}$$

3.3 Can PCA be used to reduce the dimensionality of a highly nonlinear dataset? Explain.

PCA isn't a great choice when it comes to reducing dimensionality of highly nonlinear datasets. PCA aims to reduce dimensionality by identifying the direction which captures the most variance in the data. However, if the data is highly nonlinear, PCA may not capture the true underlying structure.

For example, if we have a spiral dataset, all of the points would end up projected on a single line, and the original structure of the data would not be there anymore.

3.4 When might be sensible to chain two different dimensionality reduction algorithms? You can support your answer with an example.

3.5 How can you assess the effectiveness of a dimensionality reduction algorithm, used as a preprocessing step, on your dataset by considering the accuracy or error of a downstream model?

We can assess the effectiveness of a dimensionality reduction algorithm, used as a preprocessing step, on a dataset in the following way. First, we train the model using the full, non dimensionality reduced data, and evalute its performance on a downstream task $\mathcal{X}$. Afterwards, we apply the dimensionality reduction technique to the data, train the model and evaluate its performance again. Finally, we compare the two performances with each other.