CSE583/EE552 Pattern Recognition and Machine Learning: Homework #1

Due on January 29, 2021 at 11:59pm

PROFESSOR Yanxi Liu Spring 21

Jackson Zaunegger **jsz5092@psu.edu**

Problem 1

I think that having symmetry makes identifying patterns very easy and simple, however I do not believe that all patterns need to have symmetry. As unpleasant as it may be, lets think about serial murderers for a moment. In law enforcement, the police will look for patterns in crimes and these patterns may present themselves in many different ways. For example the police may look at a killers motivation, their victimology, and the actual nature of the crime itself. These components of a criminal investigation do not have any sort of visual symmetry, but they are still in fact patterns.

In terms of a problem we can actually classify with machine learning, I still do not believe that symmetry is necessary. If we wanted to create a image classification system to classify fruit, the images would not have symmetry. Apples, oranges, and bananas have no sort of symmetry. A banana may appear to have bilateral symmetry, but they are not perfect symmetric, much like most naturally occurring things. But just because these examples do not have symmetry does not mean they do not share patterns. A pattern with oranges for example is that they are generally spherical in shape, they are various shades of orange, and have a bumpy texture to them. Using these sort of visual features, we can still sucessfully perform classification. To formally answer the question, I do not believe that having symmetry is a necessary condition for finding a pattern.

Problem 2

The original form of the error function is

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2 \tag{1}$$

which can then be re-written into its matrix form of

$$E(w) = \frac{1}{2}(XW - T)^T(XW - T) \tag{2}$$

which we can then take the derivative from resulting in

$$\frac{\partial}{\partial W} = X^T(XW - T) \tag{3}$$

After finding w^* we can infer the polynomial $y(x, w^*)$

$$y(x, w^*) = w_0^* + w_1^*x + w_2^*x^2 + ... + w_M^*x^M = \sum_{j=0}^{M} w_j x^i \tag{4}$$

Problem 3

1. I found the following definition on Springer Link, [2] *"The curse of dimensionality, first introduced by Bellman [1], indicates that the number of samples needed to estimate an arbitrary function with a given level of accuracy grows exponentially with respect to the number of input variables (i.e., dimensionality) of the function."* Essentially what this means is that as the dimensional of the data grows larger, the volume of space that the data fills becomes less and less filled. This higher dimensionality also will increase the amount of processing time, as it will become more difficult to draw conclusions from this higher dimensional data.

Problem 3 continued on next page. . .

2. I think that this idea of the "Curse of Dimensionality" is a double-edged sword. On one hand, it will drastically increase the amount of computational resources required if you want to work with large, or more complex data samples. For example when performing image classification, the size of the images will matter drastically. As image size increases, so does the number of dimensions. This is not ideal, as the amount of processing will increase as well. On the other hand, if you have a larger dataset as whole you will have more room to add more samples into this dimensional space. For example the imagenet dataset that has millions and millions of images, would not perform well in a lower dimensional space, as the available volume in this data space will fill very quickly, where that is not as much of a problem in higher dimensional spaces. So I think it can be a curse if working will smaller simpler datasets, but could turn out to be a blessing if the dataset you are using is very large.

Problem 4

The central limit theorem is a mathematical concept that states that the distribution of a sample mean becomes more normal as the sample size increases. This will happen regardless of the populations distribution. Another interesting component is as the sample size increases, the mean will not change but the standard deviation will lessen as that sample size increases. This definition was gathered from sources [3] and [4].

1. If we were to plot out this distribution on a grid, where the X axis represents the probability of flipping a head or tails(where we have two columns, one for heads and another for tails), we should expect the distribution to be slightly skewed to whichever direction heads is. This is because the probability of flipping a head is slightly larger than flipping a tail. If these probabilities were equal, we would expect that the distribution would normalize.

2. As Central Limit Theorem states a distribution will tend to normalize as the sample size increases. In this case if we consider our sample size to now be 5000 instead of just 5, this distribution should become normal.

References

[1] Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

[2] Chen L. (2009). *Curse of Dimensionality*. springer.

[3] Ganti Akhilesh.(2019) *Central Limit Theorem*. Investopedia. `https://www.investopedia.com/terms/c/central_limit_theorem.asp` Accessed: 01-27-2021

[4] Khan S. *Central limit theorem*. Khan Academy. `https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/sampling-distribution-mean/v/central-limit-theorem` Accessed: 01-27-2020