Machine Learning, Spring 2018: Exercise Sheet 2 – with partial solution

Problem 1 (a conceptual challenge, not a mathematical challenge). Consider the set of all photographic images that have been posted on the web, and assume they have been normalized to a size of 1000 by 1000 pixels. Color intensity values are between 0 (black) and 1 (white). An image thus normalized therefore is a point in the three-million-dimensional unit hypercube $I = [0\ 1]^{3,000,000}$ (the factor 3 comes from the three RGB color intensities needed to define each pixel). But most points in I, if you would print their corresponding image, will not look like a photograph of some real-world scene at all – they will just be "grayish-tinted noise" pictures. The real world and the physics of taking photographic images impose many constraints on points in I – only points in I which satisfy these constraints will appear like a photograph of a real-world scene. Mathematically speaking, these constraints confine the "real photo-like" points to a relatively low-dimensional manifold R within I. Your task: develop an educated guess on what the dimension of this manifold R. – This is not an easy problem, because you have to think about all kinds of regularities and factors which govern the appearance of real-world-like images. Some of these regularities will be derived from low-level geometry (like "neighboring pixels have a tendency to have similar color values"), others will be derived from just how our world is ("if there are two nearby patches that appear like eyes, then very likely there is a patch in the image close by that looks like a nose"). Any such regularity trims down the three million dimensions of I toward the dimension of R. In summary, how low-dimensional do you think is the "Flickr data manifold"?

Solution. There is no known "correct" solution to this problem. One difficulty in coming up with a reasonable number lies in the circumstance that the "Flickr data manifold" is a noisy affair – real-world images to not lie on a "clean" R-dimensional manifold, but are spread randomly in a 3,000,000 dimensional "hull" around an R-dimensional manifold where they are "concentrated" (compared Figure 5B in the LN: the shaded area around the 1-dimensional line manifold has the full "thickness" dimension of the embedding space, which is 2 in that simple example). So the question of determining R really is the question of deciding how many of the 3,000,000 dimensions used to "hull up" the lower-dimensional picture manifold can be ignored, that is, considered as meaningless noise. This question has no clean answer. For instance, in a picture showing the foliage of a tree, where one leaf in the picture exhibits a single brown pixel among many green pixels – is that brown pixel noise, or might it be the photographic reflection of a grain of dirt on the leaf?

Given such considerations, one possible approach to find a meaningful number for R would go like this: First, think of how you would describe *verbally* what you can see on a 1000 by 1000 pixel image. Let us assume that with a 1-page plain English description per image you can provide a characterization that is detailed enough for your purposes to capture what is shown on the image. A 1-page plain English text has about 3000 characters, taken from an alphabet of size 30 (let's say). Each character has a bit information content of $\log_2(30) \approx 5$. The information content of a 3000 character text is thus 5 times 3000 ≈ 15000. In fact it is less than that because of inherent constraints in English texts – estimates of the average information content in English state that in such texts, each character carries on average only 1 bit. The information content of a 1-page text would thus only be about 3000 bits. Coding this by the location of a binary "one hot" vector would need 3000 dimensions. Seen in this way, one would boldly state that $R = 3000$. As it happens, this is the same ballpark as the $R = 4096$ decision that was adopted by the TICS designers.

Problem 2 Consider the set $G = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, y = x^2\}$. This is the set of all points belonging to the *graph* of the square function within the unit square. It is a 1-dimensional manifold, of the familiar parabola shape, embedded in the 2-dimensional interval $[0, 1] \times [0, 1]$. Give a map $f: G \to \mathbb{R}$, which "flattens" the parabola, similar to what you see is happening in panel **a** in Figure 4 in the lecture notes. Specifically, f should preserve distance along the parabola. That is, if two points (x, y), (x', y') have a distance d, measured along the parabola, then the distance between $f(x, y)$ and $f(x', y')$ should also be d.

Hint: this problem leads to evaluating an integral. At https://en.wikipedia.org/wiki/List_of_integrals_of_irrational_functions you will find the integral formula that solves the integral that you will get.

Solution. This boils down to compute the length of the parabola line between two points on it. Let $x' > x$. Then the length of the parabola line segment between (x, x^2) and (x', x'^2) is the integral

$$\int_x^{x'} \sqrt{1+4t^2}\,dt = \left(x' \sqrt{\frac{1}{4} + x'^2} + \frac{1}{4} \ln(x' + \sqrt{\frac{1}{4} + x'^2}\,) - x \sqrt{\frac{1}{4} + x^2} - \frac{1}{4} \ln(x + \sqrt{\frac{1}{4} + x^2}\,) \right) =: g(x, x')$$

(I looked up this formula for the integral on https://en.wikipedia.org/wiki/List_of_integrals_of_irrational_functions).

Specifically, $g(0, x) = x \sqrt{\frac{1}{4} + x^2} + \frac{1}{4} \ln\left(x + \sqrt{\frac{1}{4} + x^2} \right) - \frac{1}{4} \ln \frac{1}{2}$, and we can put

$f((x, y)) = g(0, x)$.