Homework-1 Solution

October 14, 2018

1 Nearest Neighbors and the Curse of Dimensionality

(a)

$$\mathbf{X}, \mathbf{Y} \sim \mathbf{U}\,[0,1]$$
$$\mathbf{Z} = (\mathbf{X} - \mathbf{Y})^2$$

Then, we have

$$\begin{aligned}
\mathbb{E}\,[\mathbf{Z}] &= \mathbb{E}\left[(\mathbf{X}-\mathbf{Y})^2\right] \\
&= \mathbb{E}\left[\mathbf{X}^2 + \mathbf{Y}^2 - 2\mathbf{X}\mathbf{Y}\right] \\
&= \mathbb{E}\left[\mathbf{X}^2\right] + \mathbb{E}\left[\mathbf{Y}^2\right] - \mathbb{E}\,[2\mathbf{X}\mathbf{Y}]
\end{aligned} \tag{1}$$

and

$$\begin{aligned}
\mathrm{Var}\,[\mathbf{Z}] &= \mathbb{E}[\mathbf{Z}^2] - \mathbb{E}[\mathbf{Z}]^2 \\
&= \mathbb{E}\left[\mathbf{X}^4 - 4\mathbf{X}^3\mathbf{Y} + 6\mathbf{X}^2\mathbf{Y}^2 - 4\mathbf{X}\mathbf{Y}^3 + \mathbf{Y}^4\right] + \mathbb{E}\left[\mathbf{X}^2 + \mathbf{Y}^2 - 2\mathbf{X}\mathbf{Y}\right]^2 \\
&= \mathbb{E}\left[\mathbf{X}^4\right] - 4\mathbb{E}\left[\mathbf{X}^3\right]\mathbb{E}\,[\mathbf{Y}] + 6\mathbb{E}\left[\mathbf{X}^2\right]\mathbb{E}\left[\mathbf{Y}^2\right] - 4\mathbb{E}\,[\mathbf{X}]\,\mathbb{E}\left[\mathbf{Y}^3\right] + \mathbb{E}\left[\mathbf{Y}^4\right] \\
&\quad + \left(\mathbb{E}\left[\mathbf{X}^2\right] + \mathbb{E}\left[\mathbf{Y}^2\right] - 2\mathbb{E}\,[\mathbf{X}]\,\mathbb{E}\,[\mathbf{Y}]\right)^2.
\end{aligned} \tag{2}$$

Given that both $\mathbf{X}$ and $\mathbf{Y}$ are sampled uniformly form unit interval, we can get the n-th moment of $\mathbf{X}$ and $\mathbf{Y}$ as follows,

$$\begin{aligned}
\mathbb{E}\,[\mathbf{X}] = \mathbb{E}\,[\mathbf{Y}] &= \int_0^1 x\;dx = \frac{1}{2} \\
\mathbb{E}\left[\mathbf{X}^2\right] = \mathbb{E}\left[\mathbf{Y}^2\right] &= \int_0^1 x^2\;dx = \frac{1}{3} \\
\mathbb{E}\left[\mathbf{X}^3\right] = \mathbb{E}\left[\mathbf{Y}^3\right] &= \int_0^1 x^3\;dx = \frac{1}{4} \\
\mathbb{E}\left[\mathbf{X}^4\right] = \mathbb{E}\left[\mathbf{Y}^4\right] &= \int_0^1 x^4\;dx = \frac{1}{5}
\end{aligned}$$

Plugging the above results of moment into Eqn. (1) and Eqn. (2), we have

$$\begin{aligned}
\mathbb{E}\,[\mathbf{Z}] &= \frac{1}{6} \\
\mathrm{Var}\,[\mathbf{Z}] &= \frac{7}{180}
\end{aligned} \tag{3}$$

(b) Since $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_d$ and $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_d$ are independently sampled from uniform distribution, we have

$$\begin{aligned}
\mathbb{E}\,[\mathbf{R}] &= \sum_1^d \mathbb{E}\,[\mathbf{Z}_i] = d\mathbb{E}\,[\mathbf{Z}] \\
\mathrm{Var}\,[\mathbf{R}] &= \sum_1^d \mathrm{Var}\,[\mathbf{Z}_i] = d\mathrm{Var}\,[\mathbf{Z}]
\end{aligned} \tag{4}$$

(c) From part **(b)**, we get the expectation and variance of $\mathbf{R}$, it is then straightforward to get the standard deviation as follows:

$$\mathrm{Std}\left[\mathbf{R}\right] = \sqrt{d}\,\mathrm{Std}\left[\mathbf{Z}\right]. \tag{5}$$

Now, we conclude that the expectation increases linearly along with the dimension d while the standard deviation increases with the rate $\sqrt{d}$. That's why in high dimensions, "most points are far away, and approximately the same distance".