HW2

Phillip Wang (pw1287)

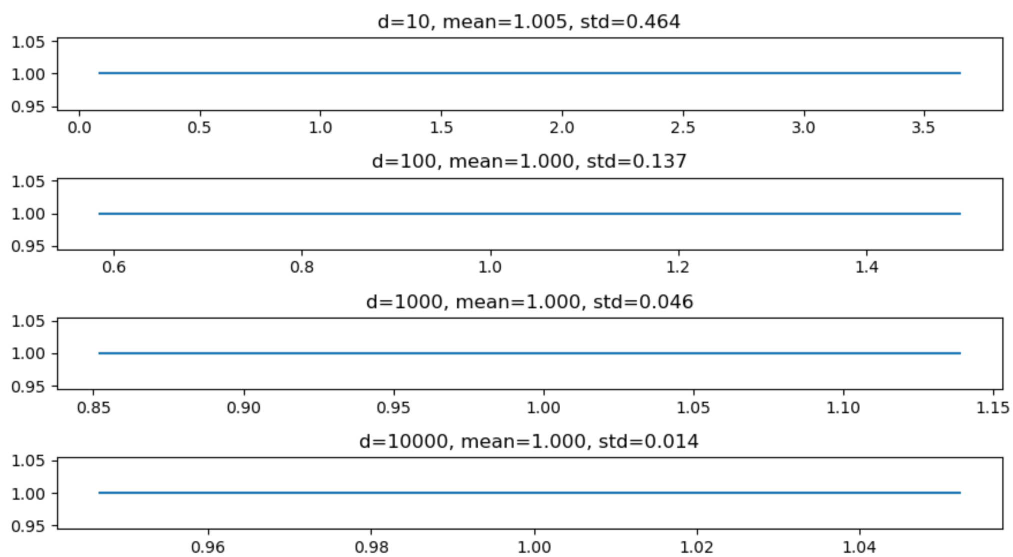
September 2021

1.

$$\mathbb{E}[\|\mathbf{X}\|^2] = \mathbb{E}[\sum_i^d x_i^2] = \sum_i^d \mathbb{E}[x_i^2] = \sum_i^d \mathbb{E}[x_i]^2 + Var[x_i] = \sum_i^d Var[x_i] = \sum_i^d \frac{1}{d} = 1$$

$$Var[\|\mathbf{X}\|^2] = Var[\sum_i^d x_i^2] = \sum_i^d Var[x_i^2] = \sum_i^d (E[x_i^4] - E[x_i^2]^2)$$
$$= \sum_i^d (\frac{3}{d^2} - \frac{1}{d^2}) = \sum_i^d (\frac{1}{d^2}) = \frac{2}{d} \implies \sigma = \frac{\sqrt{2}}{\sqrt{d}} \propto \frac{1}{\sqrt{d}}$$

2.



3.

Since $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I}/d)$, then it has a pdf of

$$f_{\mathbf{X}}([x_1, x_2 \dots x_n]) = \frac{1}{(2\pi)^{d/2}(\frac{1}{d^d})^{1/2}} \exp\left(-\frac{\mathbf{X}^T\Sigma^{-1}\mathbf{X}}{2}\right) \text{ where } \Sigma = det[\mathbf{I}/d]$$

On the other hand, by the same pdf., we have

$$\begin{aligned}\mathbb{P}[\tilde{\mathbf{X}}] = \mathbb{P}[\mathbf{R}\mathbf{X}] \propto \exp\left(-\mathbf{X}^T R^T \Sigma^{-1} R\mathbf{X}/2\right) &= \exp\left(-\mathbf{X}^T R^T d\mathbf{I} R\mathbf{X}/2\right) \\ = \exp\left(-d\mathbf{X}^T R^T R\mathbf{X}/2\right) &= \exp\left(-\mathbf{X}^T\mathbf{X}/2\right) \qquad (1) \\ &\implies \mathbb{P}[\tilde{\mathbf{X}}] \propto \mathbb{P}[\mathbf{X}]\end{aligned}$$

4.

$$\mathbb{E}[\langle \mathbf{X}, \mathbf{X}' \rangle] = \mathbb{E}[\sum_i^d x_i \cdot x_i'] = \sum_i^d \mathbb{E}[x_i \cdot x_i'] = \sum_i^d \mathbb{E}[x_i]\mathbb{E}[x_i'] = 0$$

$$\begin{aligned}Var[\langle \mathbf{X}, \mathbf{X}' \rangle] = Var[\sum_i^d x_i \cdot x_i'] = \sum_i^d Var[x_i \cdot x_i'] &= \sum_i^d \mathbb{E}[x_i {x_i'}^2] - \mathbb{E}[x_i x_i']^2 \\ = \sum_i^d Var[x_i]Var[x_i'] + Var[x_i]\mathbb{E}[x_i']^2 + Var[x_i']\mathbb{E}[x^i]^2 \\ = \sum_i^d \frac{1}{d^2} = \frac{1}{d} \implies \sigma = \frac{1}{\sqrt{d}}\end{aligned}$$

$$\|\mathbf{X} - \mathbf{X}'\| = \sqrt{\sum_i^d (x_i - x_i')^2} = \sqrt{\sum_i^d x_i^2 + \sum_i^d {x_i'}^2 - 2\sum_i^d x_i x_i'} =$$

We also have $\mathbb{E}[x_i^2] = \mathbb{E}[{x_i'}^2] = \frac{1}{d}$. So we have

$$\mathbb{E}[\sum_i^d x_i^2 + \sum_i^d {x_i'}^2 - 2\sum_i^d x_i x_i'] = \sum_i^d \mathbb{E}[x_i^2] + \sum_i^d \mathbb{E}[{x_i'}^2] - 2\sum_i^d \mathbb{E}[x_i x_i'] = 1 + 1 - 0 = 2 \implies \mathbb{E}[\|\mathbf{X} - \mathbf{X}'\|] = \sqrt{2}$$

$$\begin{aligned}Var[\sum_i^d x_i^2 + \sum_i^d {x_i'}^2 - 2\sum_i^d x_i x_i'] &= \sum_i^d Var[x_i^2] + \sum_i^d Var[{x_i'}^2] + 2\sum_i^d Var[x_i x_i'] \\ &\propto 1/d + 1/d^2 \propto 1/d \text{ for large enough } d. \\ &\implies Var[\|\mathbf{X} - \mathbf{X}'\|] =\end{aligned}$$

Hence, $\mathbb{E}[\|\mathbf{X} - \mathbf{X}'\|] = \sqrt{2}$ and $\sigma[\|\mathbf{X} - \mathbf{X}'\|] \propto \frac{1}{\sqrt{d}} \implies \|\mathbf{X} - \mathbf{X}'\| \in (\sqrt{2} \pm C/\sqrt{d})$ for large enough d.

5.

Let us assume that $\exists\, x' \in x_{1\ldots n}$ $s.t.$ $\|x' - x\|$ is minimized. Therefore, we have $\hat{f}_{NN}(x) = y_{x'} = f^*(x')$.

$$
\begin{aligned}
&\mathbb{E}[|\hat{f}_{NN}(x) - f^*(x)|] = \mathbb{E}[|f^*(x') - f^*(x)|] \\
&\quad \leq \mathbb{E}[\beta\|x' - x\|] = \beta\mathbb{E}[\|x - x'\|] \\
&\quad \text{since } \|x' - x\| \text{ is minimized, we have} \\
&\quad = \beta\mathbb{E}_{min_i}[\|x - x_i\|]
\end{aligned}
$$

6.

Let us denote $X_{(1)} = min(X_1, X_2, \ldots, X_n)$, $Y_{(1)} = min(Y_1, Y_2, \ldots, Y_n)$

$$
\begin{aligned}
\mathbb{P}(\min_i Y_i \leq t) = 1 - \mathbb{P}(\min_i Y_i > t) = 1 - (\mathbb{P}(Y_1 > t)\mathbb{P}(Y_2 > t)\ldots\mathbb{P}(Y_n > t)) \\
= 1 - \prod_{i=1}^{n}\mathbb{P}(Y_i > t) = 1 - \prod_{i=1}^{n}(1 - \mathbb{P}(Y_i \leq t)) = 1 - (1 - \mathbb{P}(Y_i \leq t))^n \\
\implies \text{pdf of } Y_{(1)} = n(1 - F(Y_i))^{n-1}f(Y_i) \\
\implies \mathbb{E}[Y_{(1)}] = \int_{-\infty}^{\infty} r \cdot n(1 - F(r))^{n-1}f(r)dr \\
\text{where } F(r), f(r) \text{ are cdf and pdf of normal distribution with } \mu = 0, \sigma = 1 \\
\text{On the other hand, } X \text{ is a random variable of } Y_{(1)}\text{. Following our previous reasoning} \\
\mathbb{E}[X_{(1)}] = \int_{-\infty}^{\infty} = t \cdot n(1 - H(t))^{n-1}h(t) \\
\text{where } H(t), h(t) \text{ are cdf and pdf of } \mathcal{N}(\mu, \sigma) \text{ respectively.} \\
\text{We also have } t = \mu + \sigma r \implies dt = \sigma dr, p_y(y) = p_x(x)|\frac{dx}{dy}| \implies h(t) = \frac{1}{\sigma}f(r), \\
F_Y(y) = y \implies H(t) = \mathbb{P}(\mu + \sigma r < t) = \mathbb{P}(r < \frac{t - \mu}{\sigma}) = F(r) \text{ where } r = \frac{t - \mu}{\sigma} \\
\mathbb{E}[X_{(1)}] = \int_{-\infty}^{\infty} (\sigma r + \mu) \cdot n(1 - F(r))^{n-1}\frac{1}{\sigma}f(r) \cdot \sigma dr
\end{aligned}
$$

$$
= \int_{-\infty}^{\infty} \sigma r n(1 - F(r))^{n-1}f(r)dr + \int_{-\infty}^{\infty} \mu n(1 - F(r))^{n-1}f(r)dr = \sigma\mathbb{E}[Y_{(1)}] + \mu\int_{-\infty}^{\infty} f(r)dr = \mu + \sigma E_n
$$

7.

By the definition of conditional independent events, each draw of x_i are i.i.d., which means $\|x - x_i\|$ are i.i.d. Therefore, we can apply our conclusion from 6., $\mathbb{E}_{min_i}\|x - x_i\| \sim \mu + \sigma E_n = \sqrt{2} + \sqrt{\frac{C}{d}}E_n$ when $d \to \infty$

8.

$$\mathbb{E}[|\hat{f}_{NN}(x) - f^*(x)|] \le \beta\mathbb{E}_{min_i}[\|x - x_i\|] = \beta(\sqrt{2} + \sqrt{\frac{C}{d}} \cdot - \sqrt{2\log(n)}) \bar{\to} \sqrt{2}\beta \text{ when } \log(n) << d$$

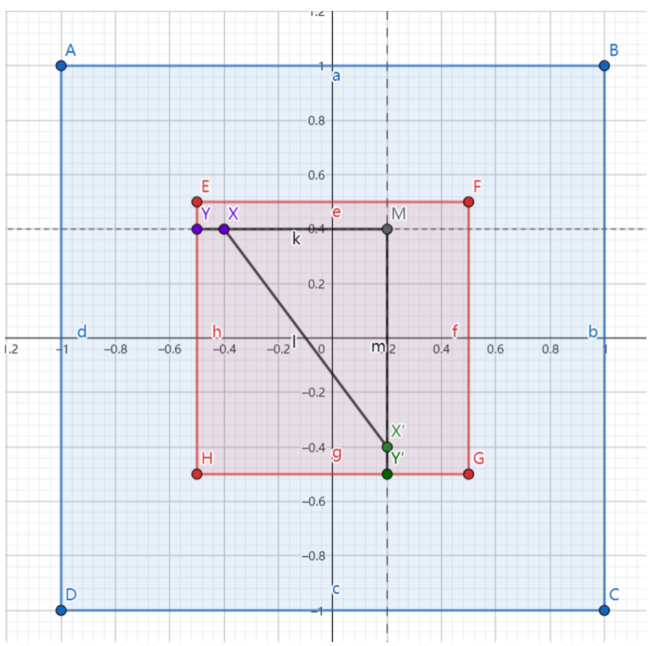
9.

Let $y_i \in \partial\Omega$ where $y_i = \underset{y_i \in \partial\Omega}{argmin} \|x_i - y_i\|$

$$\begin{aligned}
|\Psi(x) - \Psi(x')| &= |\|y - x\| - \|y' - x'\|| \\
&= |(\|y - m\| - \|x - m\|) - (\|y' - m\| - \|x' - m\|)| \\
&= | - ((\|x - m\| - \|x' - m\|) - (\|y - m\| - \|y' - m\|))| \\
&= |\|x - m\| - \|x' - m\| - (\|y - m\| - \|y' - m\|)|
\end{aligned}$$

Since we have $\|x - m\| - \|x' - m\| \le \|x - x'\|$ and $\|x - m\| - \|x' - m\| = \|x - x'\|$ iff. in a degenerate triangle where y and y' are on the same edge of the square, where we can assume without loss of generality that $\|y' - m\| = 0$. If we consider such case, we have

$$|\Psi(x) - \Psi(x')| = |\|x - x'\| - \|y - m\|| \le \|x - x'\| \implies \Psi \text{ is 1-Lipschitz}$$



10.

The verification is trivial in that $\mathcal{B}$ is capable of containing 2 non-overlapping, separable Ω in along each of its dimension, which gives us a subset with cardinality of 2^d.

11.

$$|f^*(x) - f^*(x')| = |\sum_z g(z)\Psi(x - z/2) - \sum_z g(z)\Psi(x' - z/2)|$$

$$= |\sum_z g(z)\Psi(x - z/2) - g(z)\Psi(x' - z/2)| = |\sum_z g(z)(\Psi(x - z/2) - \Psi(x' - z/2))|$$

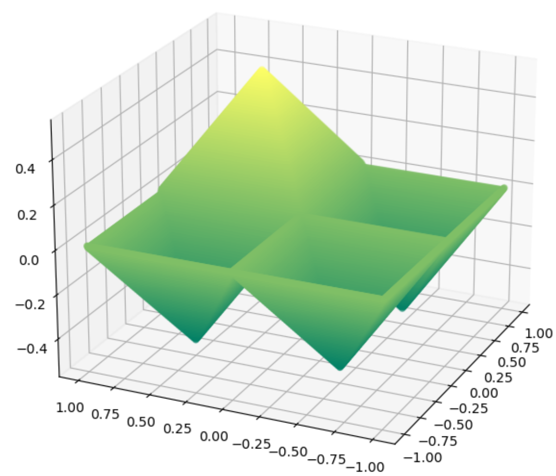
$$\leq |\sum_z \Psi(x - z/2) - \Psi(x' - z/2)|$$

We know from 9 that Ψ is 1-Lipschitz

$$\leq |\sum_z \|x - z/2 - x' + z/2\|| \leq |\|x - x'\|| \leq \|x - x'\| \implies f^* \text{ is 1-Lipschitz}$$

12.

$$g(z) = \begin{cases} 1 & \text{if } x = (1,1) \\ -1 & \text{otherwise} \end{cases}$$



13.

Since x is drawn uniformly from $[-1,1]^d$, the number of point in each tile is $\frac{n}{2^d}$. When $n \leq 2^{d-1}$, then the point in each tile $\leq \frac{1}{2}$. In other words, less or equal to half of the tiles would have data points in them. The learning algorithm could not make any meaningful inference for data in any tile without previous points. In such case, the learning algorithm could only make a random guess of

$f^*(x) = 0$ such that $|f^*(x) - \hat{f}(x)| \leq \frac{1}{2}$.

$$\frac{\mathbb{E}[|f^*(x) - \hat{f}(x)|]}{\mathbb{E}[|f^*(x)|]} = \frac{\frac{1}{n}\sum_i^n |f^*(x) - \hat{f}(x)|}{\frac{1}{n}\sum_i^n |f^*(x)|} = \frac{\sum_i^n |f^*(x) - \hat{f}(x)|}{\sum_i^n |f^*(x)|}$$
$$\geq \frac{\sum_i^n |f^*(x)| - \sum_i^n |\hat{f}(x)|}{\sum_i^n |f^*(x)|}$$

14.