Deep Learning HW2

lc3919

October 2021

Problem 1

$$Var(X_i) = \frac{1}{d} \implies E[X_i^2] = 1/d \implies E[\|\mathbf{X}\|^2] = \sum_{i=1}^{d} E[X_i^2] = 1$$

Re-writing $\|\mathbf{X}\|^2$ as $\frac{1}{d}\|\mathbf{X}\|^2$ we can now use the central limit theorem to find the distribution of $\|\mathbf{X}\|^2$

$$\sqrt{d}(\frac{1}{d}\|\mathbf{X}\|^2 - \frac{1}{d}) \to \mathcal{N}(0, \frac{1}{d})$$

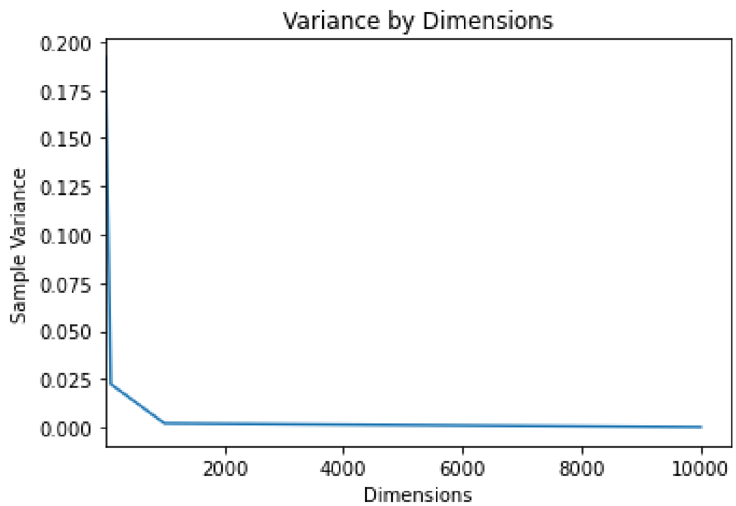
$$(\frac{1}{d}\|\mathbf{X}\|^2 - \frac{1}{d}) \to \mathcal{N}(0, \frac{1}{d^2})$$

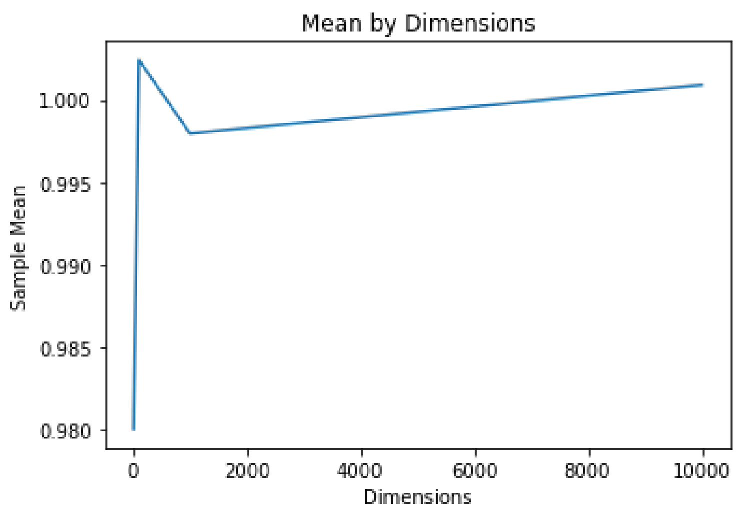
$$(\|\mathbf{X}\|^2 - 1) \to \mathcal{N}(0, \frac{1}{d})$$

$$\|\mathbf{X}\|^2 \to \mathcal{N}(1, \frac{1}{d})$$

Problem 2

Our sample variance and mean are shown to approach their asymptotic values as we increase the dimensions.





Problem 3

$$RX \sim \mathcal{N}(\mathbf{R0}, \frac{1}{d}\mathbf{RIR^T}) = \mathcal{N}(\mathbf{0}, \frac{1}{d}\mathbf{RR^TI}) = \mathcal{N}(\mathbf{0}, \frac{1}{d}\mathbf{I})$$

Problem 5

$$Var(X_iX_j) = E[(X_iX_j)^2] - E[X_iX_j]^2$$

$$Var(X_iX_j) = E[(X_iX_j)^2] - E[X_i][X_j]^2$$

$$Var(X_i X_j) = E[(X_i)^2 (X_j)^2] = E[(X_i)^2]E[(X_j)^2] = \frac{1}{d^2}$$

With the variance now known and the expectation being obviously 0 we can employ the central limit theroem.

$$\sqrt{d}(\frac{1}{d}\langle \mathbf{X}, \mathbf{X}' \rangle) \to \mathcal{N}(0, \frac{1}{d^2})$$

$$(\frac{1}{d}\langle \mathbf{X}, \mathbf{X}' \rangle) \to \mathcal{N}(0, \frac{1}{d^3})$$

$$\langle \mathbf{X}, \mathbf{X}' \rangle \to \mathcal{N}(0, \frac{1}{d})$$

Which gives us that our random variable has mean zero and standard deviation proportional to $\frac{1}{\sqrt{d}}$ Meaning that all points are nearly orthogonal as $d \to \infty$

Finally

$$\|\mathbf{X} - \mathbf{X}'\| = \|\mathbf{X}\| + \mathbf{X}'\| - 2\langle \mathbf{X}, \mathbf{X}' \rangle = \|\mathbf{X}\| + \mathbf{X}'\| \sim$$

Problem 5

$$E[|\hat{f}_{NN}(x) - f^*(x)|] = E[|\hat{f}_{NN}(x) - f^*(x_i) + f^*(x_i) - f^*(x)|] \leq E[|\hat{f}_{NN}(x) - f^*(x_i)|] + E[|f^*(x_i) - f^*(x)|]$$

We know that $\hat{f}_{NN}(x)$ assigns our training point to the cluster of the nearest element letting x_i be that element we get

$$E[|\hat{f}_{NN}(x) - f^*(x_i)] = E[|f(x_i) - f^*(x_i)] = 0$$

Then we are left with,

$$E[|\hat{f}_{NN}(x) - f^*(x)|] \leq \beta \|x - x'\|$$

Since f^* is β-lipschitz.

Problem 6

$$P(min_i X_i \leq \alpha) = 1 - \Pi_{i=1}^n P(X_i \geq \alpha) = 1 - \Pi_{i=1}^n P(\sigma Y_i + \mu \geq \alpha) = 1 - \Pi_{i=1}^n P(Y_i \geq \frac{\alpha - \mu}{\sigma})$$

$$\implies P(min_i X_i \leq \alpha) = P(min_i Y_i \leq \frac{\alpha - \mu}{\sigma})$$

Then,

$$E[min_i X_i] = E[\mu + \sigma min_i X_i] = \mu + \sigma E_n$$

Problem 7

Since $\|x - x_i\|$ and $\|x - x_j\|$ are conditionally independent given x, and they are $\mathcal{N}(\sqrt{2}, \frac{c}{d})$ random variables we can perform the same manipulations and get, That $min_i Y_i$ and $min_i \|x - x_i\|$ have the same distribution shifted and scaled the mean and standard deviation. Then we can write,

$$E[min_i \|x - x_j\|] = \sqrt{2} + \frac{\sqrt{C}}{\sqrt{d}} E_n$$

Problem 8

$$\beta(\sqrt{2} - \frac{\sqrt{C}}{\sqrt{d}} \sqrt{2\log n}) \leq \alpha$$
$$\implies \log n \geq \frac{(\frac{\alpha}{\beta} - \sqrt{2})^2 \frac{d}{c}}{2}$$
$$\implies n \geq \exp(\frac{(\frac{\alpha}{\beta} - \sqrt{2})^2 \frac{d}{c}}{2})$$

Thus showing that samples needed to provide a meaningful upper bound grows exponentially with d.

Problem 9

Let y be such that it satisfies $min_{y \in \partial\Omega} \|x - y\|$ Then for arbitrary x'

$$\|x - y\| = \|x - x' + x' - y\| \leq \|x - x'\| + \|x' - y\|$$
$$\implies \|x - y\| - \|x' - y\| \leq \|x - x'\|$$

Since $\|x - y\|$ is minimal we know for y' satisfying $min_{y \in \partial\Omega} \|x' - y'\|$

$$|\|x - y\| - \|x' - y'\|| \leq |\|x - y\| - \|x' - y\|| \leq \|x - x'\|$$

and therefore

$$|\psi(x) - \psi(x')| \leq \|x - x'\|$$

and is 1-Lipschitz.

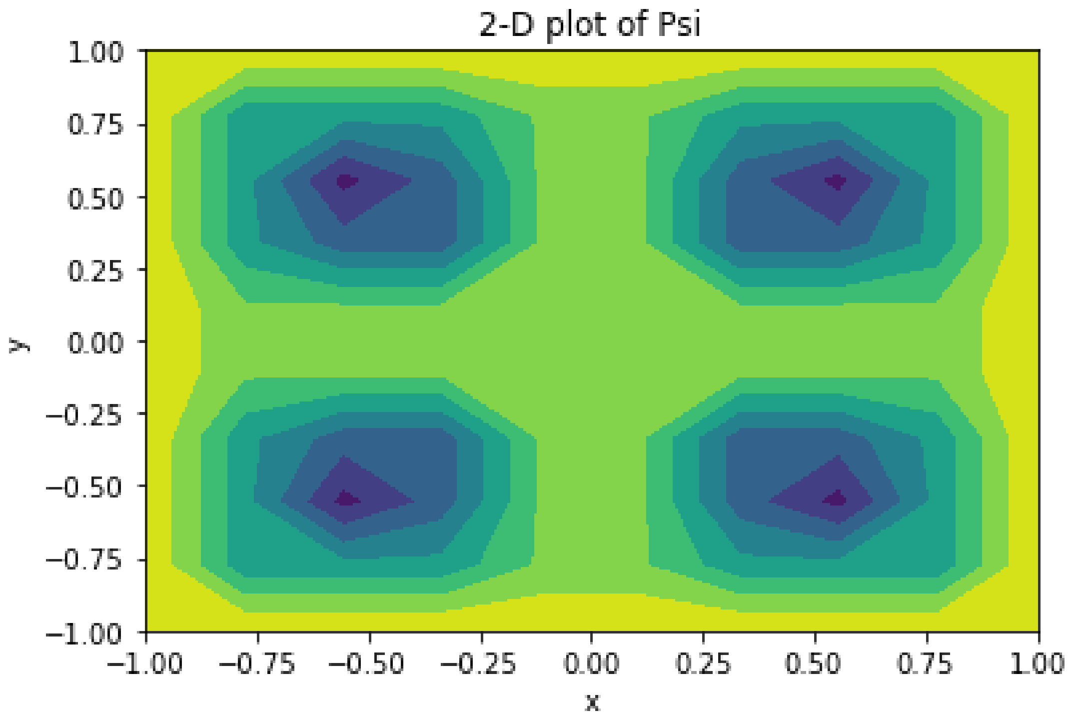
Problem 10

We can divide B into 2^d disjoint Ω d-cubes by noting that we can represent an Ω as the volume bounded by d, d-dimensional vectors. $\begin{pmatrix} \pm 1/2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \pm 1/2 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \pm 1/2 \end{pmatrix}$.

There are then 2^d of these d-cube representations corresponding to whether a 1/2 is poisive or negative.

Problem 11

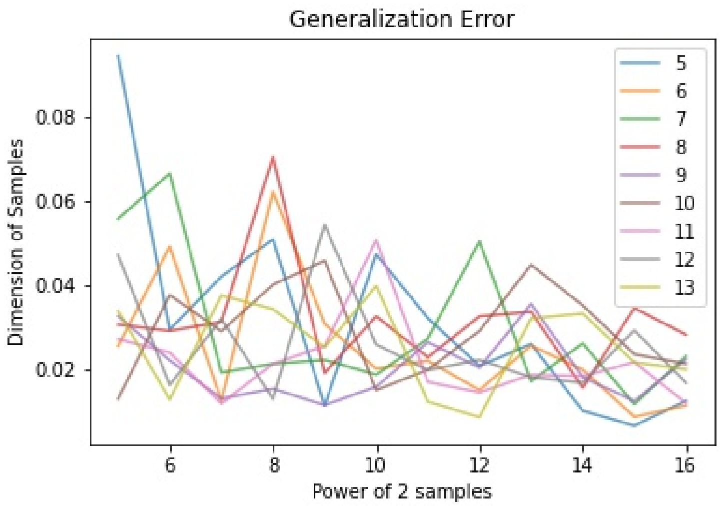
Problem 12



Problem 13

Problem 14

This roughly agrees with what we would expect to happen. In the lower dimensions there seems to be a more clear relationship between the sample size threshold of n = dimensions of input.


Generalization Error
Dimension of Samples
Power of 2 samples
0.02
0.04
0.06
0.08
6
8
10
12
14
16
5
6
7
8
9
10
11
12
13