TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik
Lehrstuhl für Datenverarbeitung
PD Dr. Martin Kleinsteuber

`Information Retrieval in High Dimensional Data`

`Assignment #1`, 03.11.2017

Due date: 16.11.2017, 6 P.M.

Please hand in your solutions via Moodle. You can add your conclusions for the PYTHON Task as comments in the PYTHON files. For the other exercises, deliver a PDF file either created using Latex or as a scan of your handwritten solution. Alternatively, you can hand in an IPython/Jupyter notebook.

Solutions can be handed in by groups of **four or five** people. Please state the group number and the names of your group members at a prominent place in your submission. (For example, at the beginning of your provided PYTHON code or in a separate text file.)

Curse of Dimensionality

Task 1: [4 Points] Let $\mathcal{C}_d = [-\frac{d}{2}, \frac{d}{2}]^p$ denote the p-dimensional hypercube of edge length d, centered at the origin and $P(A)$ denote the probability of event A.

- Assume X to be uniformly distributed in $\mathcal{C}_1$. Determine d in dependence of p, s.t.
$$P(X \in \mathcal{C}_d) = q$$
for an arbitrary $q \in [0, 1]$.
- Let the components of the p-dimensional random variable X^p be independent and have the standard normal distribution. It is known that $P(|X^1| \leq 2.576) = 0.99$. For an arbitrary p, determine the probability $P(X^p \notin \mathcal{C}_{5.152})$ for any of the components of X^p to lie outside of the interval $[-2.576, 2.576]$. Evaluate the value for $p = 2$, $p = 3$ and $p = 500$.

Task 2: [6 Points] Provide the PYTHON code to the following tasks (the code needs to be commented properly):

- Sample 100 uniformly distributed random vectors from the hypercube $[-1, 1]^p$ for $p = 2$.
- For each of the 100 vectors determine the minimum angle to all other vectors. Then compute the average of these minimum angles. Note that for two vectors

$\mathbf{x}$, $\mathbf{y}$ the cosine of the angle between the two vectors is defined as

$$\cos\left(\angle(\mathbf{x},\mathbf{y})\right) = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

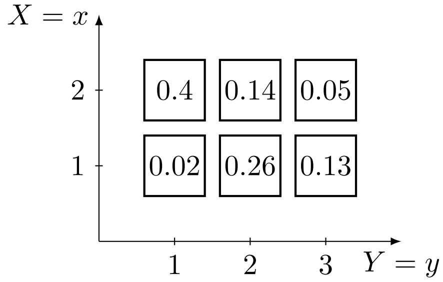
- Repeat the above for dimensions $p = 1, \dots, 1000$ and use the results to plot the average minimum angle against the dimension.
- Give an interpretation of the result. What conclusions can you draw for 2 randomly sampled vectors in a p-dimensional space?
- Does the result change if the sample size increases?

Task 3: [4 Points] Imagine a unit square with corners at $(\frac{1}{2}, \frac{1}{2}), (-\frac{1}{2}, \frac{1}{2}), (-\frac{1}{2}, -\frac{1}{2})$, and $(\frac{1}{2}, -\frac{1}{2})$.

- Draw a circle with radius $\frac{1}{2}$ around each corner (note that each circle touches its two neighboring circles). Now draw a circle around the origin with a radius such that it touches all of the four previously drawn circles. What radius does it have? Motivate your claim.
- What is the radius of the sphere around the origin constructed as above for higher dimensions (i.e. p-dimensional unit hypercube with corners at $\{-\frac{1}{2}, \frac{1}{2}\}^p \in \mathbb{R}^p$ and p-dimensional spheres with radius $\frac{1}{2}$ at each corner.) What happens when the dimension p increases towards infinity? Describe in particular what occurs for $p = 4$ (and dimensions higher than 4) and $p = 9$ (and dimensions higher than 9).

Statistical Decision Making

Figure 1: Task 3



Task 4: [6 Points] Answer the following questions. All answers must be justified.

- The numbers in Figure 1 describe the probability of the respective events (e.g. $P(X = 1,\ Y = 1) = 0.02$). Is this table a probability table? Justify your answer.
- By means of Figure 1, provide the conditional expectation $\mathbb{E}_{Y|X=2}[Y]$ and the probability of the event $X = 1$ under the condition that $Y = 3$.

- Is the function $p(x, y)$ given by

$$p(x,y) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1,\ 0 \leq y \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

  a joint density function for two random variables?

- For two random variables X and Y, let the joint density function be given by

$$p(x,y) = \begin{cases} 2e^{-(x+y)} & \text{for } 0 \leq x \leq y,\ 0 \leq y \\ 0 & \text{otherwise.} \end{cases}$$

  What are the marginal density functions for X and Y respectively?

- Let the joint density function of two random variables X and Y be given by

$$p(x,y) = \begin{cases} \frac{1}{15}(2x+4y) & \text{for } 0 < x < 3,\ 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

  Determine the probability for $X \leq 2$ under the condition that $Y = \frac{1}{2}$.