Vorlesung 20 Topologische Datenanalyse

Das Ziel von topologischer Datenanalyse (TDA) ist die **Topologie** von Datenmengen zu lernen.

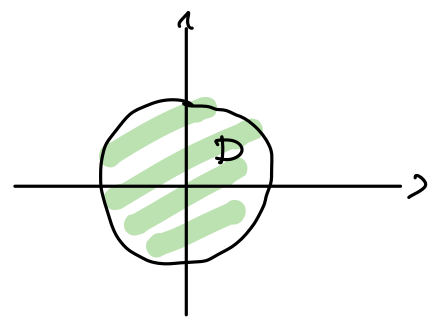
wir nennen es **Modell**

Wir nehmen an, dass die Daten von einem unbekannten Objekt in $\mathbb{R}^D$ gesampled wurden (und zwar mit Noise).

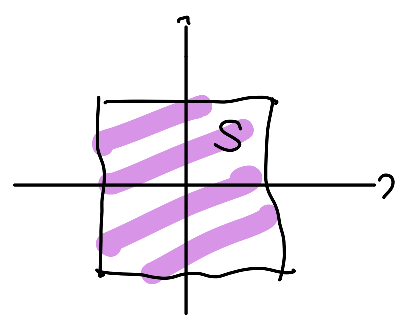
Die Topologie eines Modells gibt die **Form** bis auf kontinuierliche Deformationen an.

Beispiel 20.1:

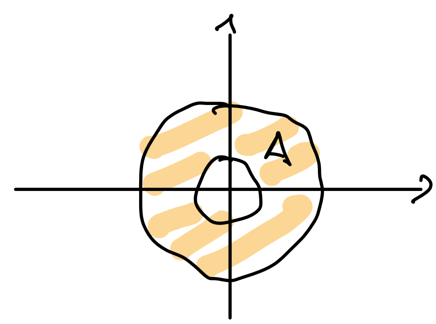
$D = \{ (x,y) \in \mathbb{R}^2 \mid x^2+y^2 \le 1 \}$



$S = \{ (x,y) \in \mathbb{R}^2 \mid \max\{|x|, |y|\} \le 1 \}$



$A = \{ (x,y) \in \mathbb{R}^2 \mid \frac{1}{2} \le x^2+y^2 \le 1 \}$



Das Ziel in TDA ist es anhand von Daten die zugrunde-liegenden Modelle zu unterscheiden, indem wir sie nach ihrer "Form" (= Topologie) klassifizieren.

Anwendungsbeispiele von TDA sind (1) das Blue Brain Projekt, wo u.A. Klassifizierung von Gehirnaktivität per TDA durchgeführt wird (2) Robotik, wo es relevant ist, ob das Modell eines Zustandsraums zusammenhängend ist.

In dieser Vorlesung wollen wir uns darauf konzentrieren die Anzahl von "Löchern" eines Modell zu lernen.

Das zugrundeliegende Konzept, um dieses Ziel zu erreichen, sind <u>simpliziale Komplexe</u>. Wir werden den Daten im $\mathbb{R}^D$ einen solchen simplizialen Komplex zuordnen und diesen dann analysieren.

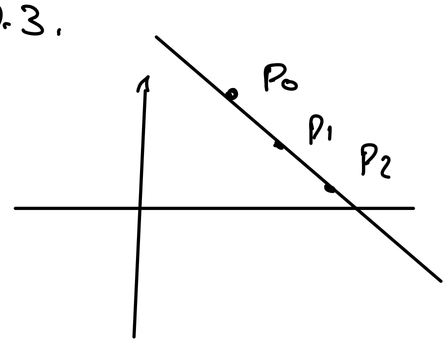
<u>Definition</u> 20.2.

Eine Mengen von Punkten $x_0, \dots, x_n \in \mathbb{R}^D$ heißen <u>affin unabhängig</u>, falls für alle $t_0, \dots, t_n \in \mathbb{R}$ mit $t_0 + \dots + t_n = 0$ gilt:

$$\sum_{i=0}^{n} t_i p_i = 0 \implies t_0 = \dots = t_n = 0.$$

D.h. $\{p_1 - p_0, \dots, p_n - p_0\}$ lin. unabh.

<u>Beispiel</u> 20.3.



Definition 20.4

Seien $P := \{p_0, \dots, p_n\} \subset \mathbb{R}^D$ affin unabh. Der n-Simplex, der von P aufgespannt wird ist

$$\Delta(P) := \{x \in \mathbb{R}^D \mid x = \sum_{i=0}^{n} t_i p_i, \sum_{i=0}^{n} t_i = 1, t_i \geq 0\}$$

Beispiel 20.5 D=3/

$n=3$