1 Алгоритм

create_embed(num):

`embed_data` = [];
for $i = 1, \ldots, num$ **do**
generate unique random neural net u
D = `get_v_D(u, embed_data)`;
append u to D and D to embed_data
`Thetas` = `poincare_embed()`
return `(embed_data, Thetas)`

2 Вспомогательные функции

Функция get_v_D(u):

create all possible neural nets with edit distance 1
by changing one element of matrix or ops of u

Фунция poincare_embed(embed_data, epochs$=10^9$**, eps**$=10^{-7}$**,lr**$=10^{-2}$**, decay**$=10^{-2}$**):**

generate vector in poincare ball for every neural net
with coordinates from uniform distribution on the interval [-0.001, 0.001]
(Thetas initialization)
while $i < epochs$ and $step\_norm > eps$ **do**
choose random neural net u
choose random neural net v with edit distance 1
choose 10 random neural nets with edit distance > 1
for every chosen neural net compute gradient
for other neural nets gradient is equal to zero vector
$step = \frac{lr}{1+decay \cdot i} \cdot gradient$
$Thetas = Thetas - step$
return $Thetas$

3 Теоретические выкладки

1. Будем минимизировать функцию $\mathcal{L}(\Theta) = \log \frac{e^{-d(u,v)}}{\sum_{v' \in \mathcal{N}(u)} e^{-d(u,v')}}$, где
   u, v - такие графы, что $edit\_distance(u, v) = 1$,
   а $\mathcal{N}(u)$ - множество, состоящее из 10 графов v', таких, что $edit\_distance(u, v') > 1$

2. Возьмем $\theta \in \Theta$, тогда:

   $d(u, v) = (1 + 2\frac{||u-v||^2}{(1-||u||^2)(1-||v||^2)})$
   $\nabla_R \mathcal{L}(\theta) = \frac{(1-\|\theta\|^2)^2}{4} \nabla_E \mathcal{L}(\theta)$

$\nabla_E \mathcal{L}(\theta) = \frac{\partial \mathcal{L}(\theta)}{\partial d(\theta,x)} \frac{\partial d(\theta,x)}{\partial \theta}$

$\frac{\partial d(\theta,x)}{\partial \theta} = \frac{4}{\beta\sqrt{\gamma^2-1}}(\frac{\|x\|^2-2\langle\theta,x\rangle+1}{\alpha^2}\theta - \frac{x}{\alpha})$, где
$\alpha = 1 - \|\theta\|^2, \beta = 1 - \|x\|^2, \gamma = 1 + \frac{2}{\alpha\beta}\|\theta - x\|^2$

3. Обозначим $\mathcal{S} = \sum_{v' \in \mathcal{N}(u)} e^{-d(u,v')}$, получим:

$\frac{\partial \mathcal{L}(u)}{\partial u} = e^{d(u,v)} \cdot \frac{-e^{-d(u,v)} \frac{\partial d(u,v)}{\partial u} \mathcal{S} - e^{-d(u,v)} \frac{\partial \mathcal{S}}{\partial u}}{\mathcal{S}} = -\frac{\partial d(u,v)}{\partial u} - \frac{\frac{\partial \mathcal{S}}{\partial u}}{\mathcal{S}}$

$\frac{\partial \mathcal{S}}{\partial u} = \sum_{v' \in \mathcal{N}(u)} \big(-e^{-d(u,v')}\big) \frac{\partial d(u,v')}{\partial u}$

$\frac{\partial \mathcal{L}(v)}{\partial v} = e^{d(u,v)} \cdot \frac{-e^{-d(u,v)} \frac{\partial d(u,v)}{\partial v} \mathcal{S}}{\mathcal{S}} = \frac{\partial d(u,v)}{\partial v}$

$\frac{\partial \mathcal{L}(v')}{\partial v'} = e^{d(u,v)} \cdot \frac{(e^{-(d(u,v)+d(u,v'))}) \frac{\partial d(u,v')}{\partial v'}}{\mathcal{S}} = \frac{e^{-d(u,v')} \frac{\partial d(u,v')}{\partial v'}}{\mathcal{S}}$