Sample Problem 2

machine learning

January 4, 2017

1 Testing of k-centers

A set $X \subset \mathbb{R}^d$ is (k, r)-clusterable if one can partition X into k clusters where each cluster's diameter is not larger than r. X is said to be ϵ-far from (k, r)-clusterable if any subset $Y \subset X, |Y| \geq (1 - \epsilon)|X|$ is not (k, r)-clusterable. Show that the presented algorithm satisfies: accept if X is (k, r)-clusterable; reject with probability at least 0.99 if X is ϵ-far from $(k, 2r)$-clusterable.

Algorithm 1. $i = 1$*; flag=true;*
let p_1 *be an arbitrary point from* X.
While $i < k + 1$ *and flag do*
(a)Uniformly choose $\ln(100k)/\epsilon$ *points from* X.
(b) If there is a sample x *such that* $\forall j \leq i, d(p_j, x) > r$ *then* $i = i + 1$*;* $p_i = x$*; continue;*
(c) flag=false;
endwhile
If $i \leq k$ *then accept else reject.*

2 VC-dimension for Dual Space

Let X be a groud set, $R \subset 2^X$ is family of sets which has VC-dimension d. We can also regard points in X as classifier for R, implemented by $x(r) = I[x \in r]$ for $x \in X, r \in R$. Prove the with dimension of X as classifiers for R is at most 2^{d+1}.

3 Sum of Bernoullis

Let $a_1, a_2, ..., a_n \in \mathbb{R}$, $X_1, X_2, ..., X_n$ are i.i.d Bernoulli ranom variable, i.e, $\mathbb{P}[X_i = \pm 1] = 1/2$. Show that there are universal constants $c, C > 0$ such that,

$$c(\sum_{i=1}^{n} a_i^2)^{1/2} \leq \mathbb{E}[|\sum_{i=1}^{n} a_i X_i|] \leq C(\sum_{i=1}^{n} a_i^2)^{1/2}$$