Machine Learning Exercises

December 24, 2016

1 K-Center

Show the K-center problem has a polynomial 2-approximation algorithm. Recall that K-center problem is the following: given a set of points $S \subset \mathbb{R}^d$ find a set of points $T \subset \mathbb{R}^d$ such that $|T| \leq k$ and $\sum_{s \in S} \min_{t \in T} d(s,t)$ is minimized over all such T. Show finding a $2-\epsilon$ approximation for any $\epsilon > 0$ is NP-hard.

2 VC-theory

Let X be a finite set, $S \subset 2^X$ is a family of subsets of X, S's VC dimension is d. Show that, with probability at least $1-\delta$, a set of independent uniform random samples from X, denoted by M(repeated elements counted twice), satisfies the following: $\forall C \in S, \left| \frac{|C \cap M|}{|M|} - \frac{|C|}{|X|} \right| \leq \epsilon$ provided $|M| \geq \frac{c}{\epsilon^2}(d \log \frac{d}{\epsilon} + \log \frac{1}{\delta})$ for some constant c.

3 Gaussian Density

Let $x_1, x_2, ..., x_n \in \mathbb{R}$ be n different real numbers. Prove that:

$$\int_{-\infty}^{\infty} \max_{i,j \in [n], j \neq i} N(x_i, (x_i - x_j)^2)(x) dx \leq O(n \log n).$$

where $N(\mu, \sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is Gaussian's density function.