作业 2 证明题部分

1. **[10 分]** 试证：对于一组实数向量数据点的集合 $S = \{x_1, x_2, ..., x_n\}$，若定义质心 c，使得平方欧几里得距离和 $\sum_{i=1}^{n} \|x_i - c\|^2$ 最小，则质心 c 就是所有点的均值 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$。

(a)若用欧氏距离，则质心为所有点的均值.

Proof: 设 x_i 坐标为 $(x_{i1}, x_{i2} \cdots x_{im})$, $i=1 \cdots n$.

c 坐标为 $(c_1 \cdots c_m)$.

$$D(c) = \sum_{i=1}^{n} \|x_i - c\|^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} (x_{ij} - c_j)^2$$

$$D'(c) = -2\sum_{j=1}^{m}\sum_{i=1}^{n} (x_{ij} - c_j)$$

$$D''(c) = 2mn > 0.$$

$\Rightarrow$ 当 $D'(c)=0$ 时 $D(c)$ 有 min.

令 $D'(c)=0$ 得 $c_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ $j = 1, 2 \cdots m$

即当 $c = \bar{x}$ 时平方欧几里德距离最小.

$\Rightarrow$ 质心为所有点的均值 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

1. **[10 分]** 试证：对于一组实数向量数据点的集合 $S = \{x_1, x_2, ..., x_n\}$，若定义质心 c，使得曼哈顿距离和 $\sum_{i=1}^{n} |x_i - c|$ 最小，则质心 c 就是所有点的中位数。

(b)若用曼哈顿距离.则质心为所有点中位数.

Proof: 设.x_i 坐标为 $(x_{i1}, x_{i2} \cdots x_{im})$ $i=1 \cdots n$.

C坐标为 $(c_1, c_2 \cdots c_m)$.

$D(c) = \sum_{i=1}^{n} |x_i - c| \geq 0$.

先看第一维的情况.

令.$d_1(c) = \sum_{i=1}^{n} |x_{i1} - c_1|$.

$\because |x_{i1} - c_1|$ 为凸函数 $\Rightarrow$.$d_1(c)$也是凸函数

$\Rightarrow$.$d_1(c)$ 有最小值存在，且取min时 $d_1'(c) = 0$.

由于 $S = \{x_1 \cdots x_n\}$ 有序性.不妨令.$x_{n1} \geq x_{21} \cdots \geq x_{n1}$.

再设.$x_{k1} \geq c_1 \geq x_{(k+1),1}$

则.$d_1(c) = \sum_{i=1}^{k} (x_{i1} - c_1) - \sum_{i=k+1}^{n} (x_{i1} - c_1)$.

$d_1'(c) = -k + (n-k) = 0$.

$k = \frac{n}{2}$

当n为偶数时.$x_{\frac{n}{2},1} \geq c_1 \geq x_{\frac{n}{2}+1,1}$ $\Rightarrow c_1$ 为该区间内任一点.

当n为奇数时.只有当 $c_1 = x_{\frac{n+1}{2}}$ 时能使 $d_1'(c) = 0$.此时 c_1 亦为中位数.

综合奇偶.只有当 c_1 为中位数时能使 $d_1(c)$ 达到最小值.即0.

同理.对于其他维度.也是 $c_2 \cdots c_m$ 取中位数时让 $d_2(c) \cdots d_m(c)$ 为0

此时 $D(c) = \sum_{j=1}^{m} d_j(c) = 0$ 取到了最小值.

$\Rightarrow$.质心 C 为所有点的中位数.