1 2.1

令 $k^* = \min_k \|t_k - \hat{y}\|$，记$\hat{y}_k = \hat{y} \cdot t_k$是$\hat{y}$第$k$位元素。不妨假设$\hat{y}$最大元素在第$k'$位处，$k' \neq k^*$，于是我们有

$$\begin{aligned}\|t_{k^*} - \hat{y}\|^2 - \|t_{k'} - \hat{y}\|^2 &= (1-\hat{y}_{k^*})^2 + \hat{y}_{k'}^2 - (1-\hat{y}_{k'})^2 - \hat{y}_{k^*}^2 \\ &= -2\hat{y}_{k^*} + 2\hat{y}_{k'}\end{aligned}$$

由于$k^* = \min_k \|t_k - \hat{y}\|$，$\|t_{k^*} - \hat{y}\|^2 - \|t_{k'} - \hat{y}\|^2 \leq 0$。又因为$\hat{y}$最大元素在第$k'$位处，所以$-2\hat{y}_{k^*} + 2\hat{y}_{k'} > 0$。

矛盾，故原假设不成立，有$k' = k^*$。

2 2.3

由题意知，最短距离的中位数d需要满足

$$p\left(\cap_{k=1}^{N}\{\|x_k\| \geq d\}\right) = \frac{1}{2}$$

由于数据点均匀分布在一个p维球内，有$p(\|x_k\| \geq d) = 1 - d^p$。考虑到每个点的位置互相独立，
$p\left(\cap_{k=1}^{N}\{\|x_k\| \geq d\}\right) = (1-d^p)^N$.

于是

$$(1-d^p)^N = \frac{1}{2}$$

$$d = \left(1 - \frac{1}{2}^{1/N}\right)^{1/p}$$

3 2.4

因为$z_i = a^T x_i$，z_i是多个标准正态分布随机变量的线性组合，所以z_i也符合正态分布。$E[z_i] = a^T E[x_i] = 0$, $\mathrm{Var}(z_i) = \|\alpha\|^2 \mathrm{Var}(x_i) = \mathrm{Var}(x_i) = 1$，所以$z_i \sim N(0,1)$。$z_i$到原点的距离平方期望为$E(z_i^2) = \mathrm{Var}(z_i) = 1$。

目标点x_t到原点的距离平方期望符合卡方分布$\mathcal{X}_p^2$，故其期望为p。

对于$p = 10$，我们有$E(\|x_t\|) = \sqrt{10} \approx 3.1$。

4 2.7

(1) 对于线性回归

$$\begin{gathered}\hat{f}(x_0) = x_0^T \hat{\beta} = x_0^T \left(X^T X\right)^{-1} X^T Y_{\hat{q}} = \textstyle\sum_{i=1}^{N} l_i(x_0, X) y_i \\ l_i(x_0, X) = x_0^T \left(X^T X\right)^{-1} x_i\end{gathered}$$

对于kNN

$$l_i(x_0, X) = \begin{cases} \frac{1}{k}, x_i \in N_k(x_0) \\ 0, \text{otherwise} \end{cases}$$

(2) 条件均方误差可以分解为

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{Y|X}}\left(f(x_0)-\hat{f}(x_0)\right)^2= \\
&\quad=\mathrm{E}_{\mathrm{Y|X}}\left[f(x_0)-\mathrm{E}_{\mathrm{Y|X}}\left[\hat{f}(x_0)\right]\right]^2+\mathrm{E}_{\mathrm{Y|X}}\left[\mathrm{E}_{\mathrm{Y|X}}\left[\hat{f}(x_0)\right]-\hat{f}(x_0)\right]^2 \\
&\quad=\mathrm{Bias}^2_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)+\mathrm{var}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)
\end{aligned}
$$

其中

$$
\begin{aligned}
&\mathrm{Bias}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)=\mathrm{E}_{\mathrm{Y|X}}\left[\mathrm{E}_{\mathrm{Y|X}}\left[\hat{f}(x_0)-f(x_0)\right]\right. \\
&\quad=\mathrm{E}_{\mathrm{Y|X}}\left[\sum_{i=1}^{N}l_i(x_0,X)y_i-f(x_0)\right]=\sum_{i=1}^{N}l_i(x_0,X)f(x_i)-f(x_0) \\
&\mathrm{Var}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)=\mathrm{Var}_{\mathrm{Y|X}}\left[\sum_{i=1}^{N}l_i(x_0,X)y_i\right]=\sigma^2\sum_{i=1}^{N}l_i^2(x_0,X)
\end{aligned}
$$

(3) 对于无条件均方误差，与上面类似

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{Y,X}}\left(f(x_0)-\hat{f}(x_0)\right)^2= \\
&\quad=\mathrm{E}_{\mathrm{Y,X}}\left[f(x_0)-\mathrm{E}_{\mathrm{Y,X}}\left[\hat{f}(x_0)\right]\right]^2+\mathrm{E}_{\mathrm{Y,X}}\left[\mathrm{E}_{\mathrm{Y,X}}\left[\hat{f}(x_0)\right]-\hat{f}(x_0)\right]^2 \\
&\quad=\mathrm{Bias}^2_{\mathrm{Y,X}}\left(\hat{f}(x_0)\right)+\mathrm{Var}_{\mathrm{Y,X}}\left(\hat{f}(x_0)\right)
\end{aligned}
$$

其中

$$
\begin{aligned}
\mathrm{Bias}_{\mathrm{X,Y}}\left(\hat{f}(x_0)\right)&=E_X\left\{\mathrm{E}_{\mathrm{Y|X}}\left[\sum_{i=1}^{N}l_i(x_0,X)y_i-f(x_0)\right]\right\} \\
&=E_X\left[\sum_{i=1}^{N}l_i(x_0,X)f(x_i)\right]-f(x_0) \\
\mathrm{Var}_{\mathrm{X,Y}}\left(\hat{f}(x_0)\right)&=E_{\mathrm{X,Y}}\left[\hat{f}(x_0)-E_{\mathrm{X,Y}}\left(\hat{f}(x_0)\right]^2=E_{\mathrm{X}}E_{\mathrm{Y|X}}\left(\hat{f}(x_0)-E_{\mathrm{X,Y}}\left(\hat{f}(x_0)\right)^2\right. \\
&=E_{\mathrm{X}}E_{\mathrm{Y|X}}\left(\sum_{i=1}^{N}l_i(x_0,X)y_i-E_{\mathrm{X}}E_{\mathrm{Y|X}}\left(\sum_{i=1}^{N}l_i(x_0,X)y_i\right)\right)^2 \\
&=E_{\mathrm{X}}\left[\sigma^2\sum_{i=1}^{N}l_i(x_0,X)^2\right]+\mathrm{Var}_{\mathrm{X}}\left[\sum_{i=1}^{N}l_i(x_0,X)f(x_i)\right]
\end{aligned}
$$

(4)

$$
\mathrm{Var}_{\mathrm{X,Y}}\left(\hat{f}(x_0)\right)=E_{\mathrm{X}}\left[\mathrm{Var}_{\mathrm{X|Y}}\left(\hat{f}(x_0)\right)\right]+\mathrm{Var}_{\mathrm{X}}\left[\sum_{i=1}^{N}l_i(x_0,X)f(x_i)\right]
$$

故无条件方差比条件方差的期望大

$$
\begin{aligned}
E_X\left[f(x_0)-\mathrm{E}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)\right]^2&=E_X\left[f(x_0)-\mathrm{E}_{\mathrm{Y,X}}\hat{f}(x_0)+\mathrm{E}_{\mathrm{Y,X}}\hat{f}(x_0)-\mathrm{E}_{\mathrm{YYX}}\left(\hat{f}(x_0)\right]^2\right) \\
&=\left[f(x_0)-\mathrm{E}_{\mathrm{Y,X}}\hat{f}(x_0)\right]^2+E_X\left[\mathrm{E}_{\mathrm{Y,X}}\hat{f}(x_0)-\mathrm{E}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)\right]^2
\end{aligned}
$$

即$E_X\left[f(x_0)-\mathrm{E}_{\mathrm{Y|X}}\left(\hat{f}(x_0)\right)\right]^2\geq\left[f(x_0)-\mathrm{E}_{\mathrm{Y,X}}\hat{f}(x_0)\right]^2$，条件偏倚均值期望大于无条件偏倚。