EPE of k-NN

$y_i \in \mathbb{R}$
$x_i \in \mathbb{R}^d$ $\quad f: \mathbb{R}^d \to \mathbb{R}$

(1) $y_i = f(x_i) + \varepsilon_i$, $\quad i = 1, \quad , n$, where

ε_i are i.i.d. and $\quad \mathbb{E}[\varepsilon_i] = 0$
$\mathbb{E}[\varepsilon_i^2] = \sigma^2 < \infty$

(2) $y = f(x) + \varepsilon$ is an unseen label for sample $x \in \mathbb{R}^d$.
$\varepsilon \in \mathbb{R}$ is indep. of $\{\varepsilon_i\}_{i=1}^n$, $\mathbb{E}[\varepsilon] = 0$, $\mathbb{E}[\varepsilon^2] = \sigma^2 < \infty$.

(3) $\hat{f}(x) = \frac{1}{k} \sum_{i \in N_k(x)} y_i$ is the k-NN estimator.

$$\boxed{\text{Claim: } \mathbb{E}\left[(y - \hat{f}(x))^2\right] = \sigma^2 + \left(f(x) - \frac{1}{k} \sum_{i \in N_k(x)} f(x_i)\right)^2 + \frac{\sigma^2}{k}}$$

Proof: $\mathbb{E}\left[(y - \hat{f}(x))^2\right] = \mathbb{E}\left[(y - f(x) + f(x) - \hat{f}(x))^2\right]$

$$= \mathbb{E}\left[(y - f(x))^2 + (f(x) - \hat{f}(x))^2 + 2 \cdot (y - f(x))(f(x) - \hat{f}(x))\right]$$

$$= \mathbb{E}\left[(y - f(x))^2\right] + \mathbb{E}\left[(f(x) - \hat{f}(x))^2\right] + 2\,\mathbb{E}\left[(y - f(x))(f(x) - \hat{f}(x))\right]$$

From (2):

$$\mathbb{E}\left[(y - f(x))^2\right] = \mathbb{E}[\varepsilon^2] = \sigma^2$$

From (3), $\hat{f}(x)$ is a function of ε_i, $i \in N_k(x)$.
From (2), y is a function of ε, which is indep.

of ε_i, $i=1,\dots,n$. Hence y, $\hat{f}(x)$ are indep.,
as functions of indep. random variables.

Thus

$$\mathbb{E}\left[(y - f(x)) \cdot (f(x) - \hat{f}(x)) \right]$$

$$\overset{\text{indep.}}{=} \mathbb{E}[\underbrace{y - f(x)}_{=\mathbb{E}[\varepsilon]=0 \text{ by } (1)}] \cdot \mathbb{E}[f(x) - \hat{f}(x)]$$

$$= 0$$

On the other hand,

$$(f(x) - \hat{f}(x))^2 = (f(x) - \mathbb{E}[\hat{f}(x)])^2 + (\mathbb{E}[\hat{f}(x)] - \hat{f}(x))^2$$
$$+ 2 \cdot (f(x) - \mathbb{E}[\hat{f}(x)])(\mathbb{E}[\hat{f}(x)] - \hat{f}(x))$$

As $f(x)$, $\mathbb{E}[\hat{f}(x)]$ are non-random,

$$\mathbb{E}\left[(f(x) - \hat{f}(x))^2\right] = (f(x) - \mathbb{E}[\hat{f}(x)])^2$$
$$+ \mathbb{E}\left[(\hat{f}(x) - \mathbb{E}(\hat{f}(x)))^2\right] + 2\,(f(x) - \mathbb{E}[\hat{f}(x)])\,\cancelto{0}{\mathbb{E}\left[\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right]}$$

Hence,

$$\mathbb{E}\left[(y - \hat{f}(x))^2\right] = \sigma^2 + \overbrace{(f(x) - \mathbb{E}[\hat{f}(x)])^2}^{\text{estimator bias}} + \overbrace{\mathbb{E}\left[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2\right]}^{\text{estimator variance}}$$

From (1), we have

$$\mathbb{E}[\hat{f}(x)] = \frac{1}{k} \sum_{i \in N_k(x)} \mathbb{E}[y_i] = \frac{1}{k} \sum_{i \in N_k(x)} f(x_i)$$

Moreover,

$$\left(\hat{f}(x) - \mathbb{E}\left[\hat{f}(x)\right]\right)^2 = \frac{1}{k^2}\left(\sum_{i \in N_k(x)} y_i - \sum_{i \in N_k(x)} f(x)\right)^2$$

$$= \frac{1}{k^2}\left(\sum_{i \in N_k(x)} (y_i - f(x_i))\right)^2 = \frac{1}{k^2}\left[\sum_{i \in N_k(x)} \varepsilon_i\right]^2$$

Thus

$$\mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}(\hat{f}(x))\right)^2\right] = \frac{1}{k^2} \operatorname{Var}\left(\sum_{i \in N_k(x)} \varepsilon_i\right)$$

$$= \frac{1}{k^2} \sum_{i \in N_k(x)} \operatorname{Var}(\varepsilon_i)$$

$$= \frac{1}{k^2} k \cdot \sigma^2 = \frac{k}{\sigma^2}$$

where here we used the fact that

$$\operatorname{Var}\left(\sum X_i\right) = \sum \operatorname{Var}(X_i)$$

when X_i are indep. random variables. $\square$