Recall that the data set at hand is composed of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ independent and identically distributed, with the same distribution as the generic pair $(X, Y)$. Recall that a local averaging estimate is an estimate of the form

$$\widehat{r}_n(\mathbf{x}) = \sum_{i=1}^{n} W_{ni}(\mathbf{x}) Y_i, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d,$$

where $(W_{n1}(\mathbf{x}), \ldots, W_{nn}(\mathbf{x}))$ is a weight vector and each $W_{ni}(\mathbf{x})$ is a Borel measurable function of $\mathbf{x}$ and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (not $Y_1, \ldots, Y_n$). Stone's theorem (1977) offers general necessary and sufficient conditions on the weights in order to guarantee universal $L^p$-consistency of local averaging estimates.

**Theorem 1.** *Consider the following five conditions:*

1. *There is constant $C$ such that, for every Borel measurable function $g : \mathbb{R}^d \to \mathbb{R}$ with $\mathbb{E}|g(\boldsymbol{X})| < \infty$,*

$$\mathbb{E}\Big[\sum_{i=1}^{n} |W_{ni}(\boldsymbol{X})||g(\boldsymbol{X}_i)|\Big] \leq C\mathbb{E}|g(\boldsymbol{X})|, \quad \textit{for all } n \geq 1.$$

2. *There is a constant $D \geq 1$ such that, for all $n \geq 1$,*

$$\sum_{i=1}^{n} |W_{ni}(\boldsymbol{X})| \leq D \quad \textit{almost surely.}$$

3. *For all $a > 0$,*

$$\sum_{i=1}^{n} |W_{ni}(\boldsymbol{X})| \mathbb{1}_{\|\boldsymbol{X}_i - \boldsymbol{X}\| > a} \to 0, \quad \textit{in probability.}$$

4. *One has*

$$\sum_{i=1}^{n} W_{ni}(\boldsymbol{X}) \to 1, \quad \textit{in probability.}$$

5. *One has*

$$\max_{1 \leq i \leq n} |W_{ni}(\boldsymbol{X})| \to 0, \quad \textit{in probability.}$$

*If $(i)-(v)$ are satisfied for any distribution of $\boldsymbol{X}$, then the corresponding regression function estimate $r_n$ is universally $L^p$-consistent ($p \geq 1$), that is,*

$$\mathbb{E}|r_n(\boldsymbol{X}) - r(\boldsymbol{X})|^p \to 0,$$

*for all distribution of $(\boldsymbol{X}, Y)$ with $\mathbb{E}|Y|^p < \infty, p \geq 1$.*