(10 pts) *1. Bias-Variance Decomposition*

State and prove the bias-variance decomposition for the predictive risk of a regression estimator $\widehat{m}_n(x)$ at x,

$$R_x(\widehat{m}_n, m) = \mathbb{E}_n \mathbb{E}(Y - \widehat{m}_n(x))^2$$

under the model $Y = m(x) + \epsilon$ where $\epsilon \sim N(0, \sigma^2)$. The expectation $\mathbb{E}$ is over the noise ϵ and the expectation $\mathbb{E}_n$ is over the training data $(X_1, Y_1), \ldots, (X_n, Y_n)$.

$$R_x(\hat{m}_n, m) = \mathbb{E}_n \mathbb{E}(Y - \hat{m}_n(x))^2 = \mathbb{E}_n \mathbb{E}(m(x) + \varepsilon - \hat{m}_n(x))^2$$

$$= \mathbb{E}_n \mathbb{E}\left[(m(x) - \hat{m}_n(x))^2 + 2\varepsilon(m(x) - \hat{m}_n(x)) + \varepsilon^2\right]$$

$$= \underbrace{\mathbb{E}_n (m(x) - \hat{m}_n(x))^2}_{\text{this is the "risk"}} + \cancel{2\overset{0}{\mathbb{E}(\varepsilon)}\, \mathbb{E}_n(m(x) - \hat{m}_n(x))} + \mathbb{E}\varepsilon^2$$

$$= \mathbb{E}_n\left[(m(x) - \mathbb{E}_n\hat{m}_n(x)) + (\mathbb{E}_n\hat{m}_n(x) - \hat{m}_n(x))\right]^2 + \sigma^2$$

$$= (m(x) - \mathbb{E}_n\hat{m}_n(x))^2 + \cancel{2(m(x) - \mathbb{E}_n\hat{m}_n(x))\,\mathbb{E}_n\left[\mathbb{E}_n\hat{m}_n(x) - \hat{m}_n(x)\right]}^{\,0} + \mathbb{E}_n\left[(\mathbb{E}_n\hat{m}_n(x) - \hat{m}_n(x))^2\right] + \sigma^2$$

$$= \underbrace{(m(x) - \mathbb{E}_n\hat{m}_n(x))^2}_{\text{bias}^2} + \underbrace{\mathbb{E}_n\left[(\mathbb{E}_n\hat{m}_n(x) - \hat{m}_n(x))^2\right]}_{\text{variance}} + \underbrace{\sigma^2}_{\text{noise}}$$

(over)

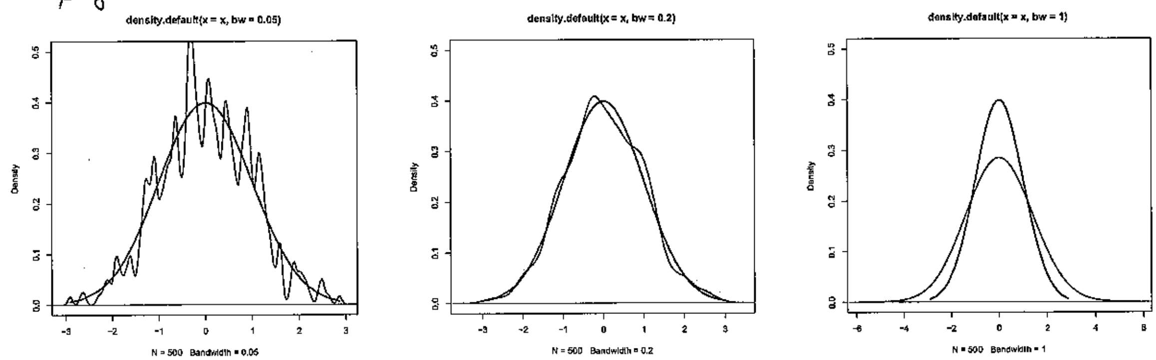
(10 pts) 2. *True or False?*

Indicate whether each of the following statements is true or false, by circling your answer.

1. Consider the three density estimates below. The true density is Gaussian, and is plotted together with the kernel density estimate $\widehat{f}_n$ for different bandwidths.

large var, small biase

small var, large bias



(i) (ii) (iii)

TRUE FALSE (a) The estimate in (ii) has a higher variance than the estimate in (i).

TRUE FALSE (b) The estimate in (iii) has a higher bias than the estimate in (ii).

2. As the bandwidth h of a local linear regression estimator increases:

TRUE FALSE (a) the variance decreases.

TRUE FALSE (b) the squared bias decreases.

TRUE FALSE (c) the risk increases.

3. Suppose that $Y_i = m(x_i) + \sigma\epsilon_i$ where $\epsilon_i \sim N(0,1)$, $x_i \in [0,1]$ and m has bounded second derivatives on $[0,1]$. Let $\widehat{m}_n(x) = \ell(x)^T Y$ be a local linear smoother with bandwidth chosen by leave-one-out cross validation. Define the confidence interval

$$I_n(x) = (\widehat{m}_n(x) - z_{\alpha/2}\sigma\|\ell(x)\|,\ \widehat{m}_n(x) + z_{\alpha/2}\sigma\|\ell(x)\|)$$

where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$. All of the following statements are with respect to the sample size n increasing to infinity.

TRUE FALSE (a) $\mathbb{P}\big(m(x) \in I_n(x)\big) \longrightarrow 1-\alpha$

TRUE FALSE (b) $\mathbb{P}\big(\mathbb{E}\widehat{m}_n(x) \in I_n(x)\big) \longrightarrow 1-\alpha$

TRUE FALSE (c) $\mathbb{P}\big(m(x) \in I_n(x)$ for all x in [0,1]$\big) \longrightarrow 1-\alpha$

TRUE FALSE (d) $\mathbb{P}\big(\mathbb{E}\widehat{m}_n(x) \in I_n(x)$ for all x in [0,1]$\big) \longrightarrow 1-\alpha$

TRUE FALSE (e) $\sup_m \mathbb{P}\big(m(x) \in I_n(x)\big) \longrightarrow 1-\alpha$

(d) is probably the trickiest one. The key is to see that $I_n(x)$ is a good CI only at fixed values of $x \in [0,1]$. To obtain confidence bands uniform in $x \in [0,1]$ one needs constructions similar to the example worked out in AoNS, §5.7.