Machine Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

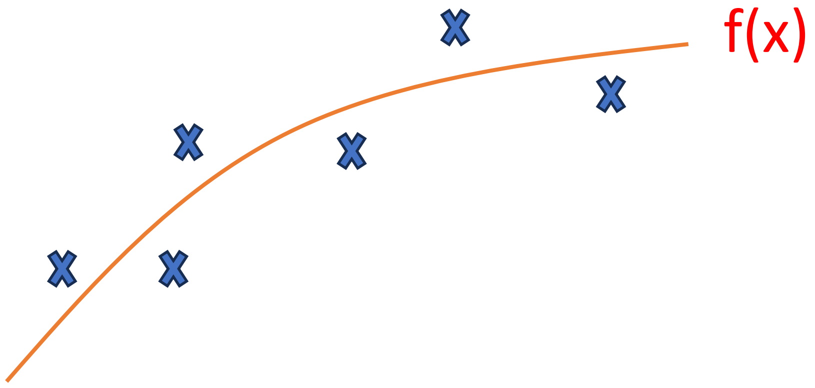
https://www.aparat.com/mehran.safayani

https://github.com/safayani/machine_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

تعریف تئوری بایاس ـ واریانس

مدل مولد داده:



$Y = f(x) + \varepsilon$

ε: نویز با توزیع D_ε که مستقل از داده ها است.

S_{train}: داده های آموزشی

D: فضای داده ها

محاسبه رابطه خطا

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}\left[\left(f(\mathbf{x})+\varepsilon-f_{S_{\text{train}}}(\mathbf{x})\right)^2\right]$$

برای یک نقطه x_0 خطا به صورت زیر است:

$$\left(f(\mathbf{x}_0)+\varepsilon-f_{S_{\text{train}}}(\mathbf{x}_0)\right)^2$$

فرض کنید که با داده های آموزشی مختلفی که از فضای داده D نمونه گیری شده اند آزمایش را تکرار میکنیم. در این حالت خطای داده x_0 به صورت زیر محاسبه می شود:

$$\mathbb{E}_{S_{\text{train}}\sim\mathcal{D},\varepsilon\sim\mathcal{D}_\varepsilon}\left[\left(f(\mathbf{x}_0)+\varepsilon-f_{S_{\text{train}}}(\mathbf{x}_0)\right)^2\right]$$

ادامه محاسبه رابطه خطا

می توانیم رابطه بالا را به صورت زیر به دست آوریم:

$$\mathbb{E}_{S_{\text{train}} \sim \mathcal{D}, \varepsilon \sim \mathcal{D}_\varepsilon}\left[\left(f(\mathbf{x}_0) + \varepsilon - f_{S_{\text{train}}}(\mathbf{x}_0)\right)^2\right]$$

$$\stackrel{(a)}{=} \mathbb{E}_{\varepsilon \sim \mathcal{D}_\varepsilon}[\varepsilon^2] + \mathbb{E}_{S_{\text{train}} \sim \mathcal{D}}\left[\left(f(\mathbf{x}_0) - f_{S_{\text{train}}}(\mathbf{x}_0)\right)^2\right]$$

$$\stackrel{(b)}{=} \mathrm{Var}_{\varepsilon \sim \mathcal{D}_\varepsilon}[\varepsilon] + \mathbb{E}_{S_{\text{train}} \sim \mathcal{D}}\left[\left(f(\mathbf{x}_0) - f_{S_{\text{train}}}(\mathbf{x}_0)\right)^2\right]$$

$$\stackrel{(c)}{=} \underbrace{\mathrm{Var}_{\varepsilon \sim \mathcal{D}_\varepsilon}[\varepsilon]}_{\text{noise variance}}$$

$$+ \underbrace{\left(f(\mathbf{x}_0) - \mathbb{E}_{S'_{\text{train}} \sim \mathcal{D}}\left[f_{S'_{\text{train}}}(\mathbf{x}_0)\right]\right)^2}_{\text{bias}}$$

$$+ \underbrace{\mathbb{E}_{S_{\text{train}} \sim \mathcal{D}}[(\mathbb{E}_{S'_{\text{train}} \sim \mathcal{D}}\left[f_{S'_{\text{train}}}(\mathbf{x}_0)\right] - f_{S_{\text{train}}}(\mathbf{x}_0))^2]}_{\text{variance}}.$$

ادامه محاسبه رابطه خطا

توجه کنید در بخش (a) عبارت زیر حذف شده است. چرا؟؟

$$\mathbb{E}_{S_{\text{train}} \sim \mathcal{D}, \varepsilon \sim \mathcal{D}_\varepsilon}\left[2\varepsilon\left(f(\mathbf{x}_0) - f_{S_{\text{train}}}(\mathbf{x}_0)\right)\right]$$

در بخش (b):

$$E_{\varepsilon \sim D_\varepsilon}[\varepsilon^2] = var_{\varepsilon \sim D_\varepsilon}[\varepsilon]$$

در بخش (c):

عبارت $\mathbb{E}_{S'_{\text{train}} \sim \mathcal{D}}\left[f_{S'_{\text{train}}}(\mathbf{x}_0)\right]$ که (S' یک مجموعه داده از D است) را به رابطه اضافه و کم می کنیم و سپس توان ۲ را اعمال می کنیم. در این رابطه یک ترم سوم هم وجود دارد که نشان می دهیم که به صورت زیر برابر با صفر است:

ادامه محاسبه رابطه خطا

$$
\begin{aligned}
& \mathbb{E}_{S\sim\mathcal{D}}\left[(f(\mathbf{x}_0) - \mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)]) \cdot \left(\mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)] - f_S(\mathbf{x}_0)\right)\right] \\
= \quad & (f(\mathbf{x}_0) - \mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)]) \cdot \mathbb{E}_{S\sim\mathcal{D}}\left[\left(\mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)] - f_S(\mathbf{x}_0)\right)\right] \\
= \quad & (f(\mathbf{x}_0) - \mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)]) \cdot (\mathbb{E}_{S'\sim\mathcal{D}}[f_{S'}(\mathbf{x}_0)] - \mathbb{E}_{S\sim\mathcal{D}}[f_S(\mathbf{x}_0)]) \\
= \quad & 0
\end{aligned}
$$

ادامه محاسبه رابطه خطا

تعبیر رابطه (c):

از سه ترم مثبت تشکیل شده است. ترم اول ربطی به نحوه آموزش مدل ندارد و ناشی از عدم قطعیت ذاتی در داده ها است.

بایاس تفاضل مابین مقدار واقعی$f(x_0)$و متوسط مدل های مختلفی است که بر روی داده ها آموزش دیده اند. (مدل های ساده نمی توانند خوب بر روی داده ها تطبیق یابند. در نتیجه بایاس زیاد می شود.)

ترم واریانس در واقع واریانس مدل های مختلفی است که آموزش دیده اند. اگر مدل ما خیلی پیچیده باشد با تغییر اندکی در داده ها شکل مدل عوض می شود و پیش بینی بر روی x_0 به میزان زیادی متغیر می شود.

$$\underbrace{\left(f(\mathbf{x}_0) - \mathbb{E}_{S'_{\text{train}} \sim \mathcal{D}}\left[f_{S'_{\text{train}}}(\mathbf{x}_0)\right]\right)^2}_{\text{bias}}$$

$$\underbrace{\mathbb{E}_{S_{\text{train}} \sim \mathcal{D}}[(\mathbb{E}_{S'_{\text{train}} \sim \mathcal{D}}\left[f_{S'_{\text{train}}}(\mathbf{x}_0)\right] - f_{S_{\text{train}}}(\mathbf{x}_0))^2]}_{\text{variance}}.$$

Examples

- $f_S(x) = k$

Prove it

Bias is high

Variance is 0

$$\boldsymbol{f_S(x) = f(x) + \varepsilon}$$

Bias is zero

Variance is $var_{\varepsilon \sim D_\varepsilon}[\varepsilon]$

چند مثال

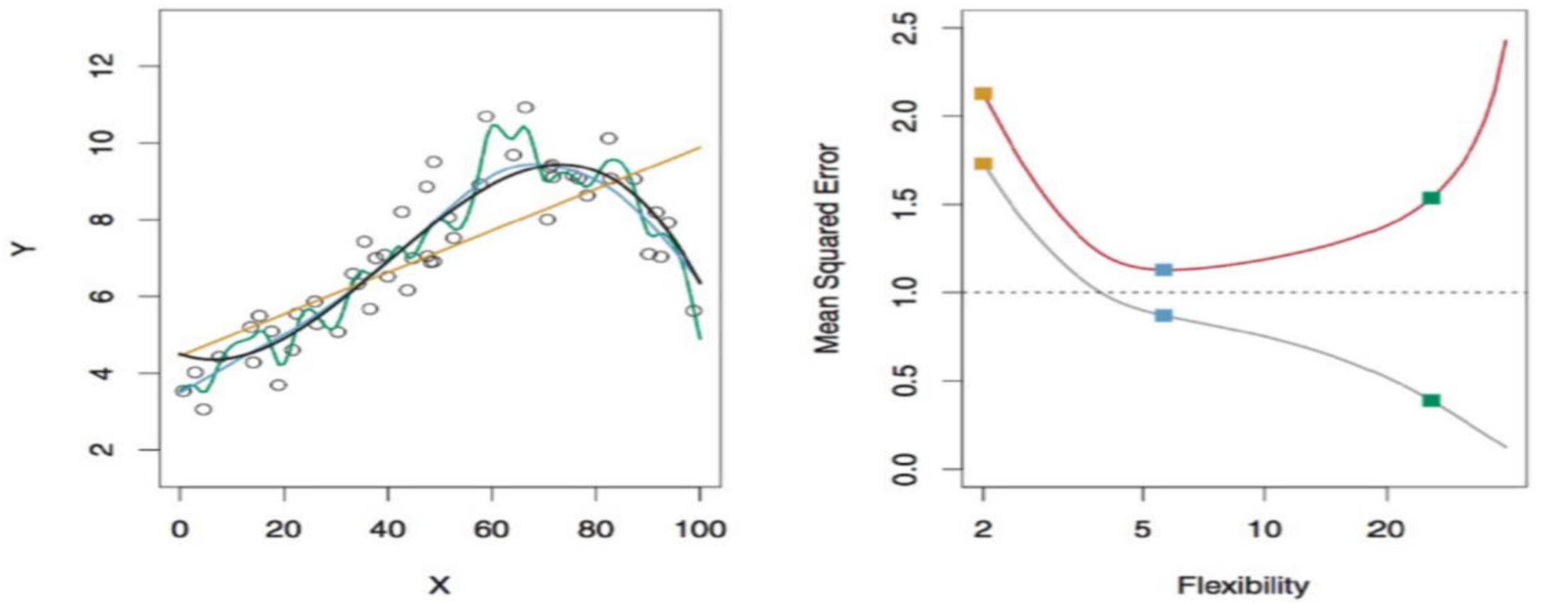


FIGURE 2.9. Left: *Data simulated from f, shown in black. Three estimates of f are shown: the linear regression line (orange curve), and two smoothing spline fits (blue and green curves).* Right: *Training MSE (grey curve), test MSE (red curve), and minimum possible test MSE over all methods (dashed line). Squares represent the training and test MSEs for the three fits shown in the left-hand panel.*

چند مثال

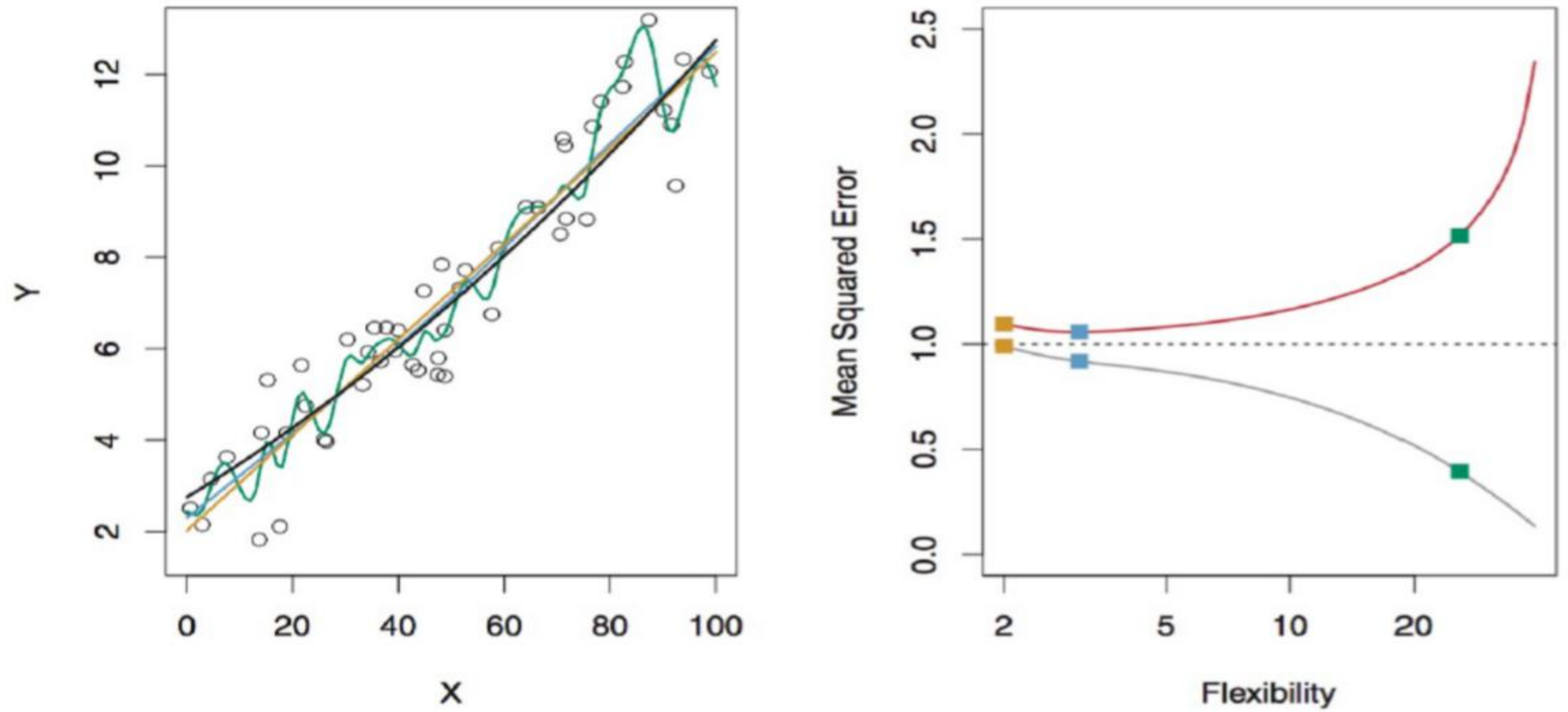


FIGURE 2.10. *Details are as in Figure 2.9, using a different true f that is much closer to linear. In this setting, linear regression provides a very good fit to the data.*

چند مثال

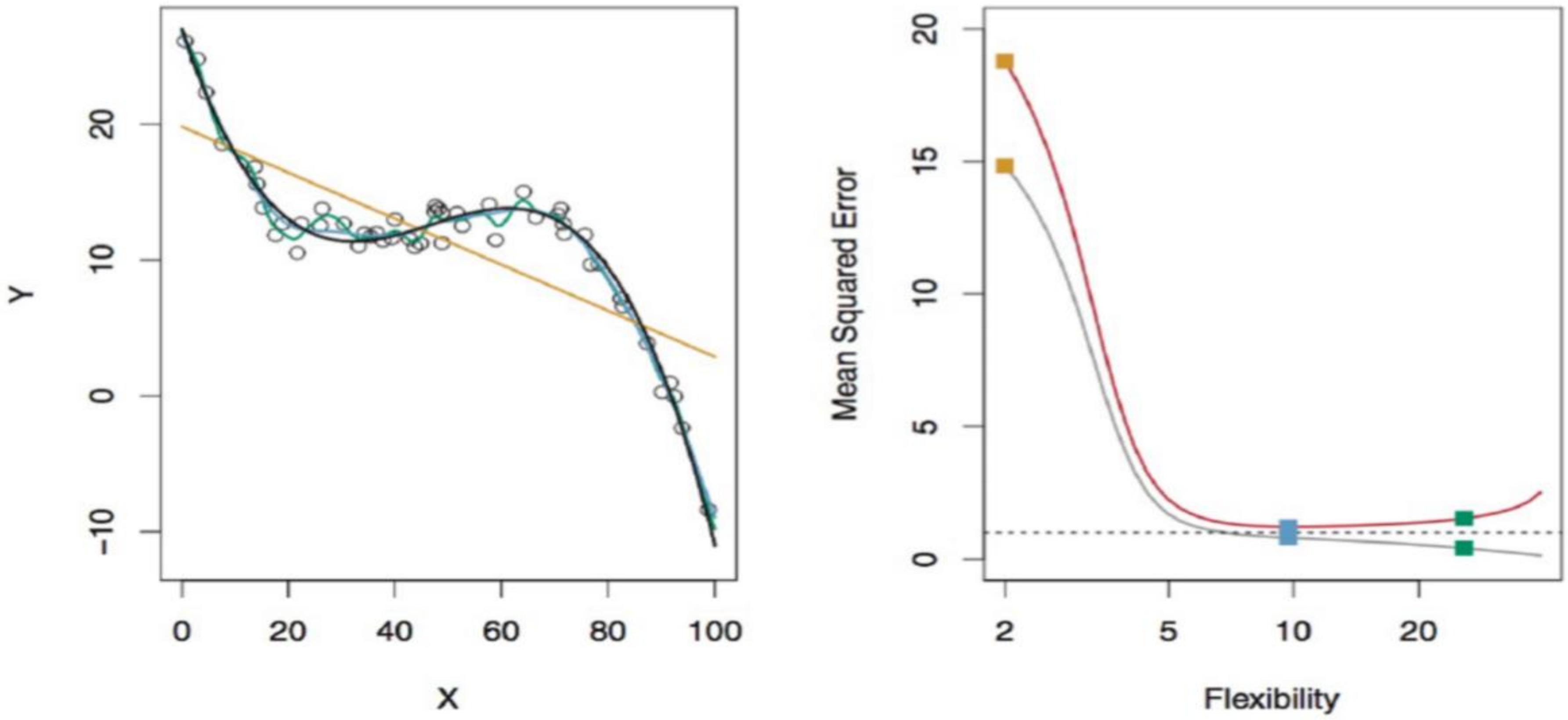


FIGURE 2.11. *Details are as in Figure 2.9, using a different f that is far from linear. In this setting, linear regression provides a very poor fit to the data.*

چند مثال

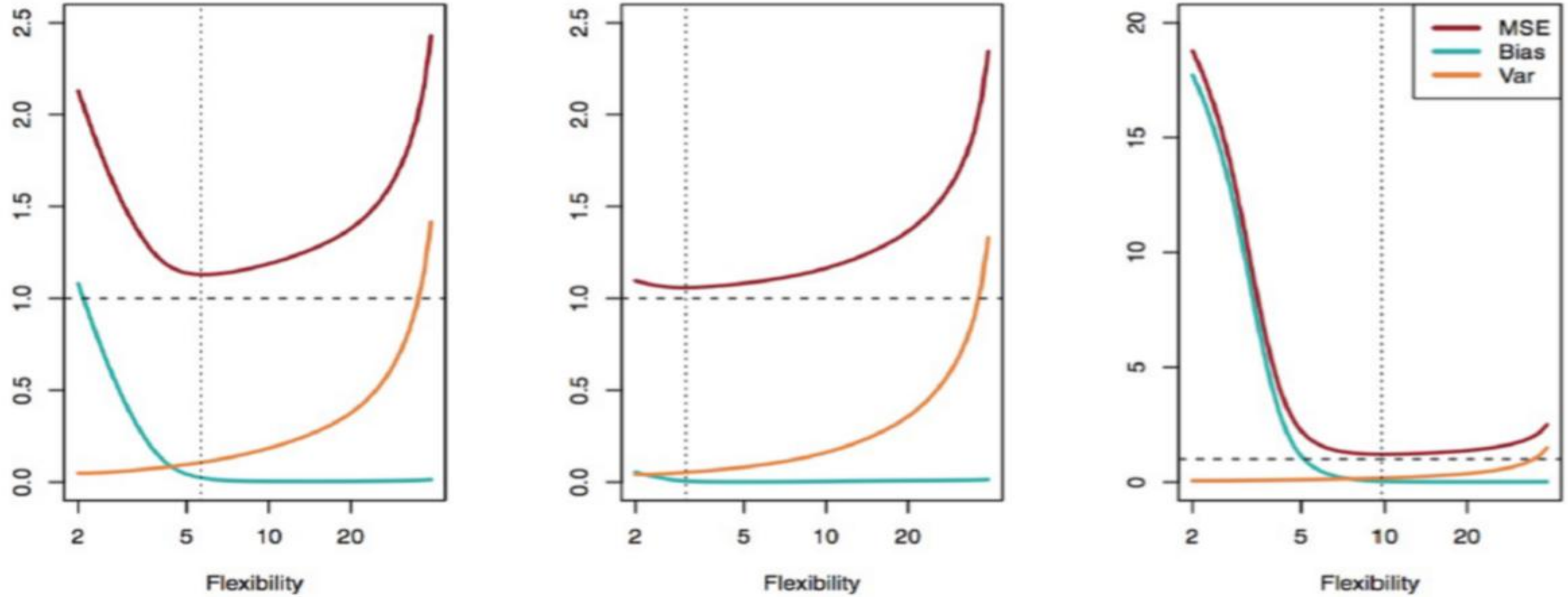


FIGURE 2.12. *Squared bias (blue curve), variance (orange curve), $Var(\epsilon)$ (dashed line), and test MSE (red curve) for the three data sets in Figures 2.9–2.11. The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.*