Aula 14/10/2020

Valor médio do MSE do conjunto de teste.

- **Training MSE versus Test MSE, em função da flexibilidade do modelo.**

- **Exercício:** estudar os três casos.

- **In practice**, one can usually compute the **training MSE** with relative ease;

- but estimating **test MSE** is considerably more **difficult** because usually **no test data** are available.

- As the previous three examples illustrate, **the flexibility level** corresponding to the model with **the minimal test MSE** can vary considerably among data sets.

- Throughout this book, we discuss **a variety of approaches** that can be used in practice to estimate this minimum point.

- One important method is **cross-validation (Chapter 5),** which is a method
for estimating test MSE using the training data.

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

Expected test MSE, for a given value x_0 : valor esperado do “set MSE”, para um valor fixo de x_0, considerando as f estimadas, a partir de vários conjuntos de treinamento. É uma média (Esperança) sobre as f estimadas.

$\hat{f}$

- É chamada de método de aprendizado estatístico.
- É uma variável aleatória que depende do conjunto de treinamento usado para estimá-la ou “fitá-la”. E assim podemos calcular sua variância.

$\mathrm{Var}(\hat{f}(x_0))$

- É como a f varia quando usamos diferentes conjuntos de treinamento.
- Característica ideal: não ser muito grande.
- In general, more flexible statistical methods have higher variance.

$\mathrm{Var}(\epsilon).$

Variância do erro não redutível.

Conjunto de treinamento: Amostra Aleatória da disciplina de Estatística.

$\boxed{\hat{f}}$

Para cada conjunto de treinamento: temos uma f estimada diferente.

Para cada conjunto de treinamento i: $X_1, X_2, X_3, \ldots, X_n$ temos uma f_i diferente.

Por exemplo, se quisermos usar uma função linear:

- $Y_i(x) = a(i,1)\, x + a(i,2) = f_i(x)$, temos para x_0 (fixo e pertencente ao conj. de teste)

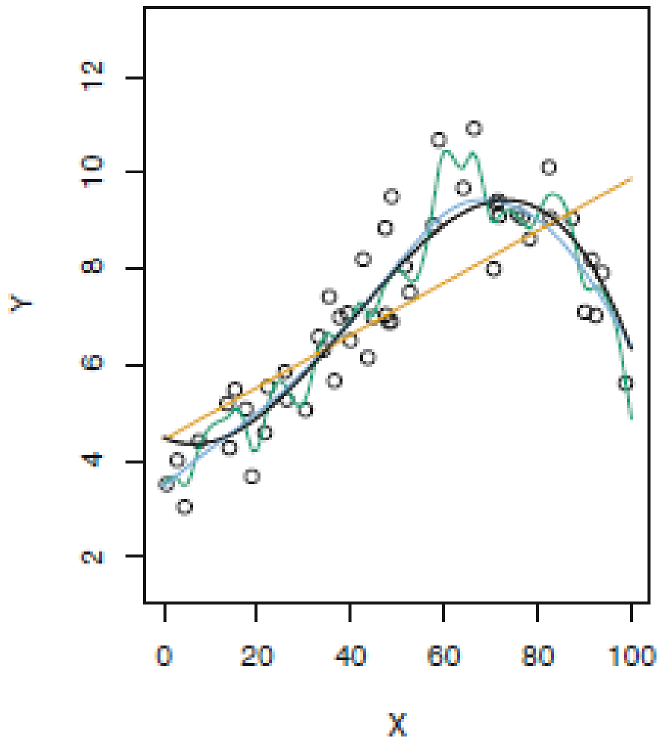
- $Y_i(x_0) = a(i,1)\, x_0 + a(i,2) = f_i(x_0)$, onde

- $a(i,1)= g_1(X_1, X_2, X_3, \ldots, X_n)$ e $a(i,2)= g_2(X_1, X_2, X_3, \ldots, X_n)$ são variáveis aleatórias.

- Como fixamos x_0 , $f_i(x_0)$ é uma variável aleatória pois é função de $a(i,1)= g_1(X_1, X_2, X_3, \ldots, X_n)$ e $a(i,2)= g_2(X_1, X_2, X_3, \ldots, X_n)$.

- Mas ainda fica a questão de qual distribuição de probabilidade que está sendo usada para calcular a Esperança.

Discutindo a Variância



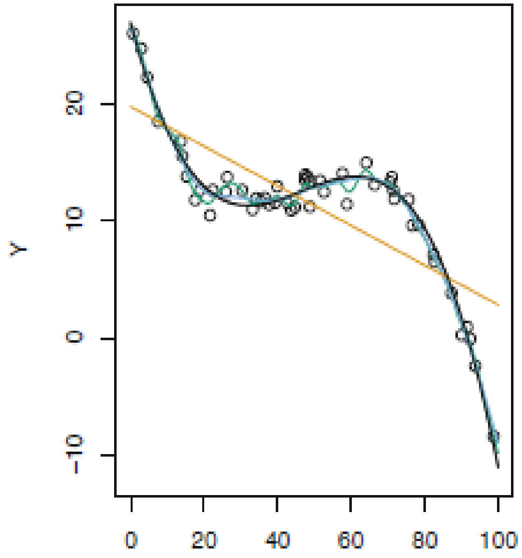
- A curva verde é flexível e se adequa bem aos dados.
- Porém tem o método tem uma variância alta: mudar alguns pontos e a curva muda.
- Em contraste, a reta laranja é bem menos flexível, mas variaria pouco ao mudarmos alguns pontos do conjunto de treinamento.

$[\text{Bias}(\hat{f}(x_0))]^2$

O viés se refere ao erros provenientes pela escolha do modelo para uma situação da vida real, que pode ser bem complicada.

Por exemplo, a regressão linear assume que há uma relação linear entre Y e $X_1, X_2, \ldots, X_p$.

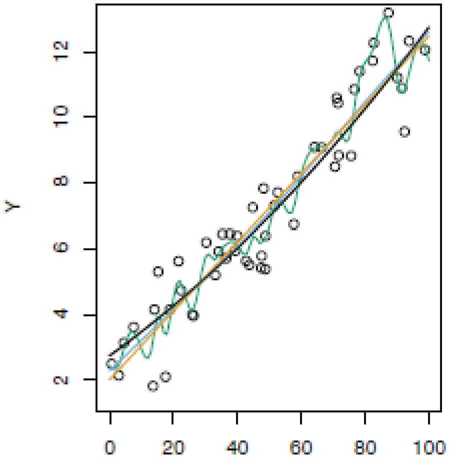
É bem pouco provável que essa relação seja linear, e uma regressão linear trará um viés para a estimação de f.



- **A f verdadeira** da Figura ao lado é explicitamente não-linear. Não adianta ter mais dados de treinamento se continuarmos usando regressão linear.

- Não tem como produzir uma estimativa mais precisa. O uso do modelo de regressão linear causa um **grande viés** para esse exemplo.

Discutindo o Viés



- Na figura ao lado a situação é bem diferente.
- A *f* verdadeira é aproximadamente linear.
- Então quanto mais dados de treinamento houver mais precisa será o modelo de regressão linear proposto.

Quanto mais flexível o modelo menos viés haverá.

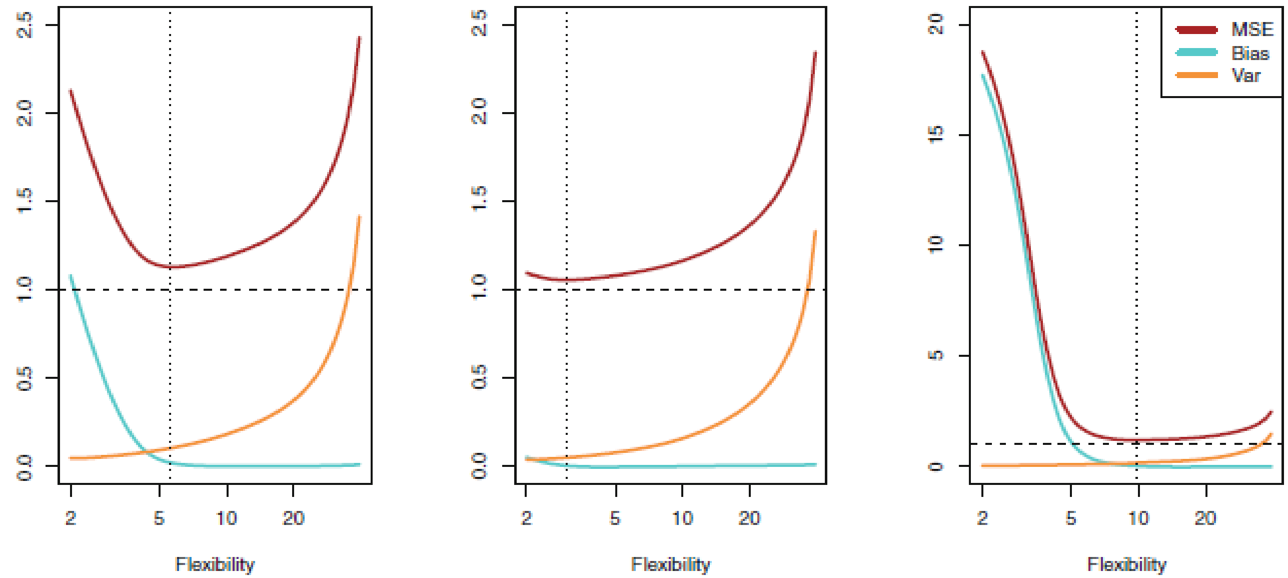
$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

E finalmente, podemos interpretar essa expressão matemática:

- Para diminuir o “test MSE”, é necessário que escolhamos um método de aprendizado estatístico que se tenha simultaneamente valores bem pequenos de variância e viés.
- Observar que a variância é sempre positiva, e que o viés está elevado ao quadrado, e portanto são ambos positivos.
- Portanto, o “test MSE” nunca vai ser menor que Var(e).

- Quanto maior a flexibilidade do método selecionado, maior a variância e menor é o viés.

- Mudanças no Viés e na Variância determinam aumento ou diminuição do test MSE médio.

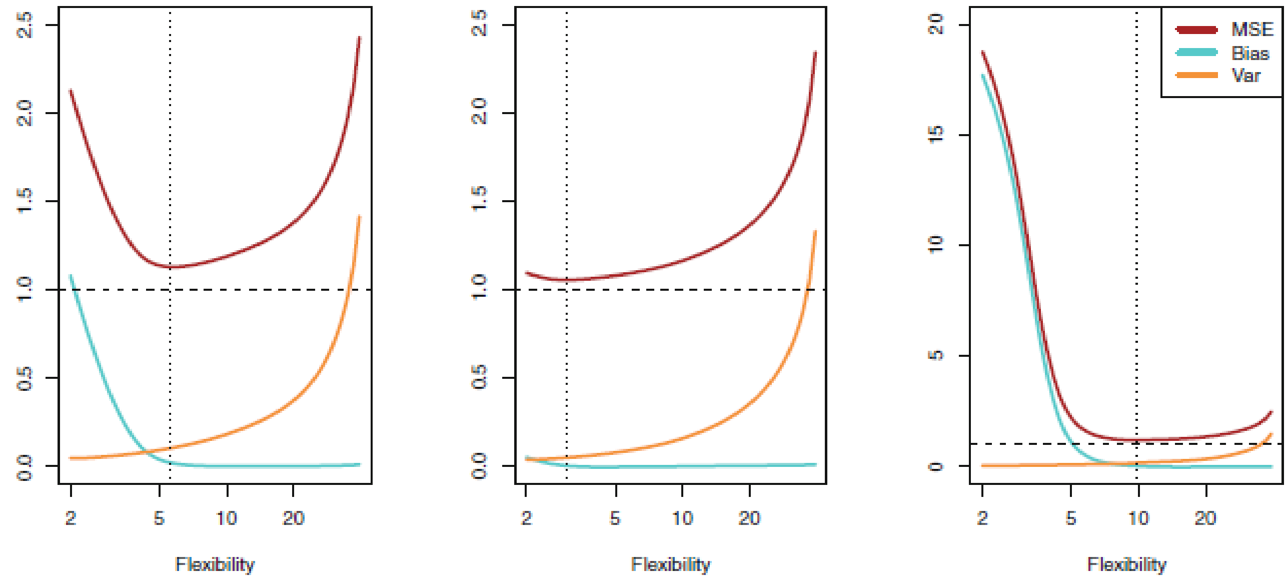


Análise dos casos já vistos, analisados segundo a expressão para o test MSE médio (em vermelho). Aparecem também a Variância (em laranja), o Viés (em azul).

- Nos 3 casos a variância aumenta quando a flexibilidade do modelo aumenta.

- A flexibilidade ótima (linha vertical tracejada) para o test MSE, para cada um dos modelos, muda consideravelmente.

Por que? Por que o viés e a variânça se modificam de forma diferente.



No primeiro gráfico, vemos que o viés decai rapidamente, resultando e um decrescimento acentuado do test MSE médio.

No segundo gráfico, a f verdadeira é linear, então o viés decai bem pouco, e o mesmo ocorre para o test MSE.

No terceiro gráfico, há um declínio violento do viés, pois a verdadeira f é não-linear. Há um pequeno crescimento da variância, então o test MSE médio também tem um declínio acentuado.

- **The relationship between bias, variance, and test set MSE** given in Equation 2.7 and displayed in Figure 2.12 is referred to as the *bias-variance trade-off (perde-ganha ou troca ou balanço)* .

Good test set performance of a statistical learning method requires:

- low variance and
- low squared bias.

This is referred to as a **trade-off** because it is easy to obtain a method with extremely low bias but high variance.

- **For instance**, by drawing a curve that passes through every single training observation

Or a method with very low variance but high bias

- **For instance** by fitting a horizontal line to the data.

The challenge lies in finding a method for which both the variance and the squared bias are low. This trade-off is one of the most important recurring themes in this book.