Bias-Variance Trade-Off

Jérémie Cabessa

Laboratoire DAVID, UVSQ

Bias-variance trade-off

- Soit une relation fonctionnelle

$$Y = f(\boldsymbol{X}) + \epsilon$$

entre des variables d'input $\boldsymbol{X} = (X_1, \ldots, X_p)$ et d'output Y. Le bruit ϵ satisfait $\mathrm{E}(\epsilon) = 0$.

- On cherche à obtenir un modèle

$$\hat{Y} = \hat{f}(\boldsymbol{X})$$

qui soit le plus performant possible *sur le test set* !

Bias-variance trade-off

- Soit une relation fonctionnelle

$$Y = f(\boldsymbol{X}) + \epsilon$$

entre des variables d'input $\boldsymbol{X} = (X_1, \ldots, X_p)$ et d'output Y. Le bruit ϵ satisfait $\mathrm{E}(\epsilon) = 0$.

- On cherche à obtenir un modèle

$$\hat{Y} = \hat{f}(\boldsymbol{X})$$

qui soit le plus performant possible *sur le test set* !

Bias-variance trade-off

- Soit un dataset

$$S = \left\{ (\boldsymbol{x_i}, y_i) \in \mathbb{R}^M \times \mathbb{R} : i = 1, \ldots, N \right\}$$

et soit $(\boldsymbol{x}, y) \in S$ un point du dataset.

- On note $\mathrm{E}_{S,\epsilon}[\ldots] := \mathrm{E}[\ldots]$.
- Par hypothèse, on a: $\mathrm{E}[\epsilon] = 0$. Et puisque la "vraie" relation fonctionnelle f est déterministe, on a: $\mathrm{E}[f] = f$.
- On rappelle que

$$\mathrm{Var}[X] := \mathrm{E}\left[(X - \mathrm{E}[X])^2 \right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 .$$

Bias-variance trade-off

- Soit un dataset

$$S = \left\{(\boldsymbol{x_i}, y_i) \in \mathbb{R}^M \times \mathbb{R} : i = 1, \dots, N\right\}$$

  et soit $(\boldsymbol{x}, y) \in S$ un point du dataset.

- On note $\mathrm{E}_{S,\epsilon}\left[\dots\right] := \mathrm{E}\left[\dots\right]$.

- Par hypothèse, on a: $\mathrm{E}\left[\epsilon\right] = 0$. Et puisque la "vraie" relation fonctionnelle f est déterministe, on a: $\mathrm{E}\left[f\right] = f$.

- On rappelle que

$$\mathrm{Var}\left[X\right] := \mathrm{E}\left[\left(X - \mathrm{E}[X]\right)^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2.$$

Bias-variance trade-off

- Soit un dataset

$$S = \left\{(\boldsymbol{x_i}, y_i) \in \mathbb{R}^M \times \mathbb{R} : i = 1, \ldots, N\right\}$$

  et soit $(\boldsymbol{x}, y) \in S$ un point du dataset.
- On note $\mathrm{E}_{S,\epsilon}[\ldots] := \mathrm{E}[\ldots]$.
- Par hypothèse, on a: $\mathrm{E}[\epsilon] = 0$. Et puisque la "vraie" relation fonctionnelle f est déterministe, on a: $\mathrm{E}[f] = f$.
- On rappelle que

$$\mathrm{Var}[X] := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2.$$

Bias-variance trade-off

- Soit un dataset

$$S = \left\{ (\boldsymbol{x_i}, y_i) \in \mathbb{R}^M \times \mathbb{R} : i = 1, \ldots, N \right\}$$

  et soit $(\boldsymbol{x}, y) \in S$ un point du dataset.
- On note $\mathrm{E}_{S,\epsilon}\left[\ldots\right] := \mathrm{E}\left[\ldots\right]$.
- Par hypothèse, on a: $\mathrm{E}\left[\epsilon\right] = 0$. Et puisque la "vraie" relation fonctionnelle f est déterministe, on a: $\mathrm{E}\left[f\right] = f$.
- On rappelle que

$$\mathrm{Var}\left[X\right] := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2.$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - y\right)^2\right] &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 - 2\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\epsilon + \epsilon^2\right] + \mathrm{E}\left[\epsilon^2\right] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{E}\left[\left(\epsilon - \mathrm{E}[\epsilon]\right)^2\right] \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 \\
&\quad - \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right] \qquad \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\right)^2\right]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - y\right)^2\right] &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 - 2\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\epsilon + \epsilon^2\right] + \mathrm{E}\left[\epsilon^2\right] \\
(\mathrm{E}(\epsilon) = 0) \quad &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{E}\left[\left(\epsilon - \mathrm{E}[\epsilon]\right)^2\right] \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 \\
&\quad - \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right] \qquad \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\right)^2\right]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - y\right)^2\right] &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 - 2\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\epsilon + \epsilon^2\right] + \mathrm{E}\left[\epsilon^2\right] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{E}\left[(\epsilon - \mathrm{E}[\epsilon])^2\right] \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 \\
&\quad - \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right] \qquad \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\right)^2\right]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - y\big)^2\Big] &= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\big)^2\Big] \\
&= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2 - 2\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)\epsilon + \epsilon^2\Big] + \mathrm{E}\Big[\epsilon^2\Big] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{E}\Big[(\epsilon - \mathrm{E}[\epsilon])^2\Big] \\
&= \mathrm{E}\Big[\hat{f}^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\Big] + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\hat{f}^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\Big] + \mathrm{E}\Big[\hat{f}(\boldsymbol{x})\Big]^2 \\
&\quad - \mathrm{E}\Big[\hat{f}(\boldsymbol{x})\Big]^2 + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{Var}\Big[\hat{f}(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big]^2 + \mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big] := \mathrm{E}\big[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big] \qquad \mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big] := \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\big)^2\Big]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - y\big)^2\Big] &= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\big)^2\Big] \\
&= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2 - 2\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)\epsilon + \epsilon^2\Big] + \mathrm{E}\Big[\epsilon^2\Big] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{E}\Big[\big(\epsilon - \mathrm{E}[\epsilon]\big)^2\Big] \\
&= \mathrm{E}\Big[\hat{f}^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat{f}(\boldsymbol{x}) f(\boldsymbol{x})\Big] + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\hat{f}^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat{f}(\boldsymbol{x}) f(\boldsymbol{x})\Big] + \mathrm{E}\Big[\hat{f}(\boldsymbol{x})\Big]^2 \\
&\quad - \mathrm{E}\Big[\hat{f}(\boldsymbol{x})\Big]^2 + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{Var}\Big[\hat{f}(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big]^2 + \mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big] := \mathrm{E}\big[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\big] \qquad \mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big] := \mathrm{E}\Big[\big(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\big)^2\Big]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - y\big)^2\Big] &= \mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\big)^2\Big] \\
&= \mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2 - 2\big(\hat f(\boldsymbol{x}) - f(\boldsymbol{x})\big)\epsilon + \epsilon^2\Big] + \mathrm{E}\Big[\epsilon^2\Big] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{E}\Big[\big(\epsilon - \mathrm{E}[\epsilon]\big)^2\Big] \\
&= \mathrm{E}\Big[\hat f^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat f(\boldsymbol{x}) f(\boldsymbol{x})\Big] + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\hat f^2(\boldsymbol{x})\Big] - 2\mathrm{E}\Big[\hat f(\boldsymbol{x}) f(\boldsymbol{x})\Big] + \mathrm{E}\Big[\hat f(\boldsymbol{x})\Big]^2 \\
&\quad - \mathrm{E}\Big[\hat f(\boldsymbol{x})\Big]^2 + \mathrm{E}\Big[f^2(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - f(\boldsymbol{x})\big)^2\Big] + \mathrm{Var}\Big[\hat f(\boldsymbol{x})\Big] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\big[\hat f(\boldsymbol{x})\big]^2 + \mathrm{Var}\big[\hat f(\boldsymbol{x})\big] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\big[\hat f(\boldsymbol{x})\big] := \mathrm{E}\big[\hat f(\boldsymbol{x}) - f(\boldsymbol{x})\big] \qquad \mathrm{Var}\big[\hat f(\boldsymbol{x})\big] := \mathrm{E}\Big[\big(\hat f(\boldsymbol{x}) - \mathrm{E}[\hat f(\boldsymbol{x})]\big)^2\Big]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - y\right)^2\right] &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 - 2\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\epsilon + \epsilon^2\right] + \mathrm{E}\left[\epsilon^2\right] \\
(\mathrm{E}(\epsilon) = 0) \qquad &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{E}\left[(\epsilon - \mathrm{E}[\epsilon])^2\right] \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 \\
&- \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$
\mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right] \qquad \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\right)^2\right]
$$

Bias-variance trade-off

$$
\begin{aligned}
\mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - y\right)^2\right] &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x}) - \epsilon\right)^2\right] \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2 - 2\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)\epsilon + \epsilon^2\right] + \mathrm{E}\left[\epsilon^2\right] \\
(\mathrm{E}(\epsilon) = 0) \quad &= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{E}\left[\left(\epsilon - \mathrm{E}[\epsilon]\right)^2\right] \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\hat{f}^2(\boldsymbol{x})\right] - 2\mathrm{E}\left[\hat{f}(\boldsymbol{x})f(\boldsymbol{x})\right] + \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 \\
&\quad - \mathrm{E}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{E}\left[f^2(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right)^2\right] + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon) \\
&= \mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right]^2 + \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] + \mathrm{Var}(\epsilon)
\end{aligned}
$$

où

$$\mathrm{Biais}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right] \quad \mathrm{Var}\left[\hat{f}(\boldsymbol{x})\right] := \mathrm{E}\left[\left(\hat{f}(\boldsymbol{x}) - \mathrm{E}[\hat{f}(\boldsymbol{x})]\right)^2\right]$$

Bias-variance trade-off

- Le biais $\mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big]$ représente **l'erreur due à la complexité du modèle $\hat{f}$.**
- La variance $\mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big]$ représente **la sensibilité du modèle $\hat{f}$ par rapport à son training set**, i.e., la variation moyenne de $\hat{f}(\boldsymbol{x})$ autour de sa moyenne $\mathrm{E}[\hat{f}(\boldsymbol{x})]$ si le modèle $\hat{f}$ était estimé à partir de différents training sets.
- $\mathrm{Var}(\epsilon)$ représente **l'erreur irréductible** liée au bruit inhérent à la vraie relation fonctionnelle f.
- **Dilemme biais-variance (bias-variance trade-off):** plus le modèle $\hat{f}$ est complexe, plus le biais sera faible, mais plus la variance sera élevée.

Bias-variance trade-off

- Le biais $\mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big]$ représente **l'erreur due à la complexité du modèle** $\hat{f}$.
- La variance $\mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big]$ représente **la sensibilité du modèle $\hat{f}$ par rapport à son training set**, i.e., la variation moyenne de $\hat{f}(\boldsymbol{x})$ autour de sa moyenne $\mathrm{E}[\hat{f}(\boldsymbol{x})]$ si le modèle $\hat{f}$ était estimé à partir de différents training sets.
- $\mathrm{Var}(\epsilon)$ représente **l'erreur irréductible** liée au bruit inhérent à la vraie relation fonctionnelle f.
- **Dilemme biais-variance (bias-variance trade-off):** plus le modèle $\hat{f}$ est complexe, plus le biais sera faible, mais plus la variance sera élevée.

Bias-variance trade-off

- Le biais $\text{Biais}\big[\hat{f}(\boldsymbol{x})\big]$ représente **l'erreur due à la complexité du modèle** $\hat{f}$.
- La variance $\text{Var}\big[\hat{f}(\boldsymbol{x})\big]$ représente **la sensibilité du modèle** $\hat{f}$ **par rapport à son training set**, i.e., la variation moyenne de $\hat{f}(\boldsymbol{x})$ autour de sa moyenne $\mathrm{E}[\hat{f}(\boldsymbol{x})]$ si le modèle $\hat{f}$ était estimé à partir de différents training sets.
- $\text{Var}(\epsilon)$ représente **l'erreur irréductible** liée au bruit inhérent à la vraie relation fonctionnelle f.
- **Dilemme biais-variance (bias-variance trade-off):** plus le modèle $\hat{f}$ est complexe, plus le biais sera faible, mais plus la variance sera élevée.

Bias-variance trade-off

- Le biais $\mathrm{Biais}\big[\hat{f}(\boldsymbol{x})\big]$ représente **l'erreur due à la complexité du modèle** $\hat{f}$.
- La variance $\mathrm{Var}\big[\hat{f}(\boldsymbol{x})\big]$ représente **la sensibilité du modèle** $\hat{f}$ **par rapport à son training set**, i.e., la variation moyenne de $\hat{f}(\boldsymbol{x})$ autour de sa moyenne $\mathrm{E}[\hat{f}(\boldsymbol{x})]$ si le modèle $\hat{f}$ était estimé à partir de différents training sets.
- $\mathrm{Var}(\epsilon)$ représente **l'erreur irréductible** liée au bruit inhérent à la vraie relation fonctionnelle f.
- **Dilemme biais-variance (bias-variance trade-off):** plus le modèle $\hat{f}$ est complexe, plus le biais sera faible, mais plus la variance sera élevée.

Bias-variance trade-off

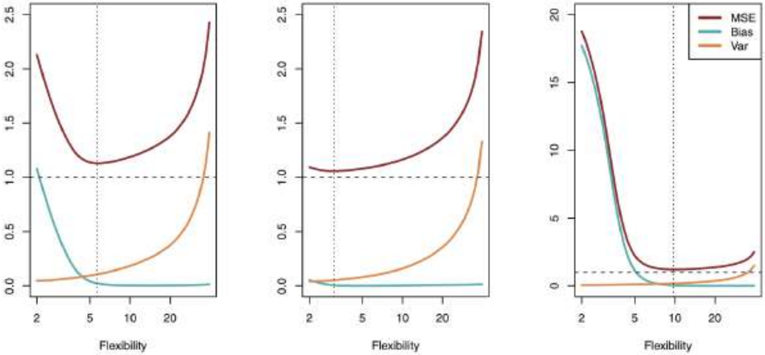

FIGURE 2.12. *Squared bias (blue curve), variance (orange curve), Var(ϵ) (dashed line), and test MSE (red curve) for the three data sets in Figures 2.9–2.11. The vertical dotted line indicates the flexibility level corresponding to the smallest test MSE.*

Figure taken from [James et al., 2013]

Bibliographie

Fleuret, F. (2022).
Deep Learning Course.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013).
An Introduction to Statistical Learning: with Applications in R, volume 103 of *Springer Texts in Statistics*.
Springer, New York.

Wikipedia contributors (2022).
Bias–variance tradeoff — Wikipedia, the free encyclopedia.