Nonlinear transforms

$$\mathbf{x} = (x_0, x_1, \cdots, x_d) \quad \xrightarrow{\Phi} \quad \mathbf{z} = (z_0, z_1, \cdots\cdots\cdots\cdots, z_{\tilde{d}})$$

Each $z_i = \phi_i(\mathbf{x})$ $\qquad \mathbf{z} = \Phi(\mathbf{x})$

Example: $\mathbf{z} = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$

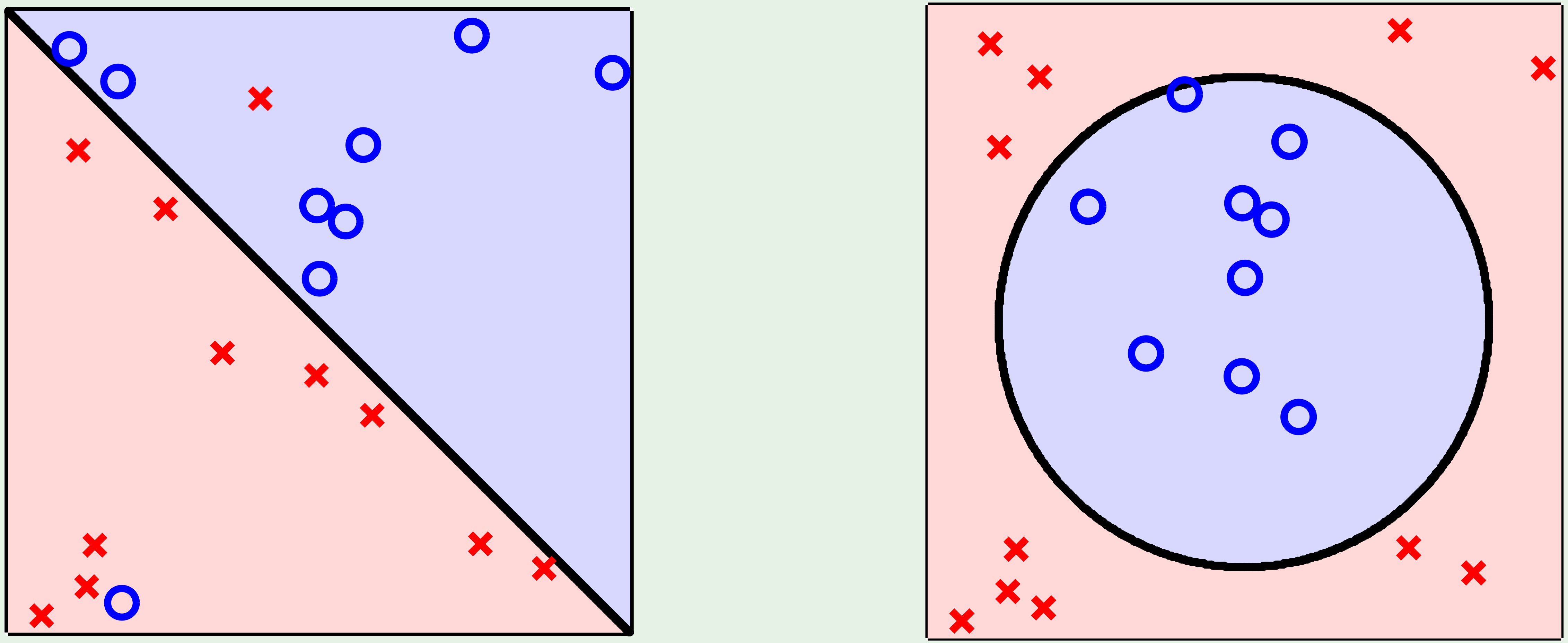
Final hypothesis $g(\mathbf{x})$ in $\mathcal{X}$ space:

$\operatorname{sign}\left(\tilde{\mathbf{w}}^{\mathsf{T}}\Phi(\mathbf{x})\right)$ or $\tilde{\mathbf{w}}^{\mathsf{T}}\Phi(\mathbf{x})$

The price we pay

$$\begin{tikzcd}
\mathbf{x} = (x_0, x_1, \cdots, x_d) \arrow[r, "\Phi"] \arrow[d] & \mathbf{z} = (z_0, z_1, \cdots\cdots\cdots\cdots, z_{\tilde{d}}) \arrow[d] \\
\mathbf{w} & \tilde{\mathbf{w}} \\
d_{\text{VC}} = d + 1 & d_{\text{VC}} \leq \tilde{d} + 1
\end{tikzcd}$$

Two non-separable cases

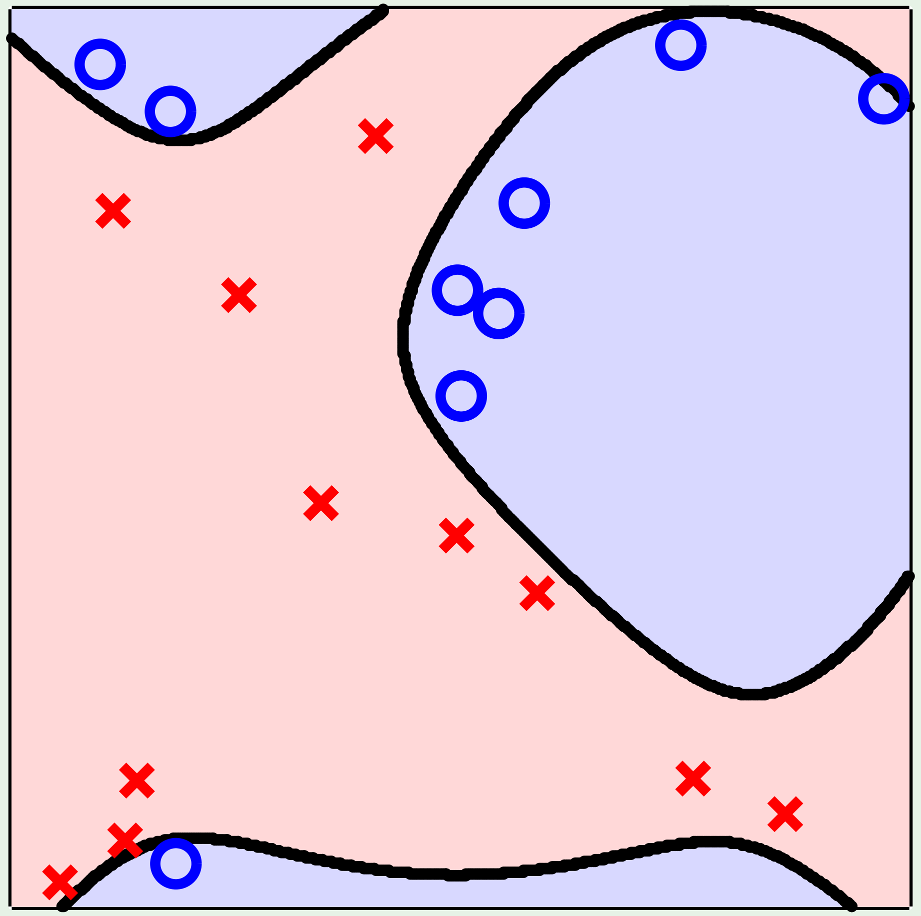


First case

Use a linear model in $\mathcal{X}$; accept $E_{\text{in}} > 0$

or

Insist on $E_{\text{in}} = 0$; go to high-dimensional $\mathcal{Z}$



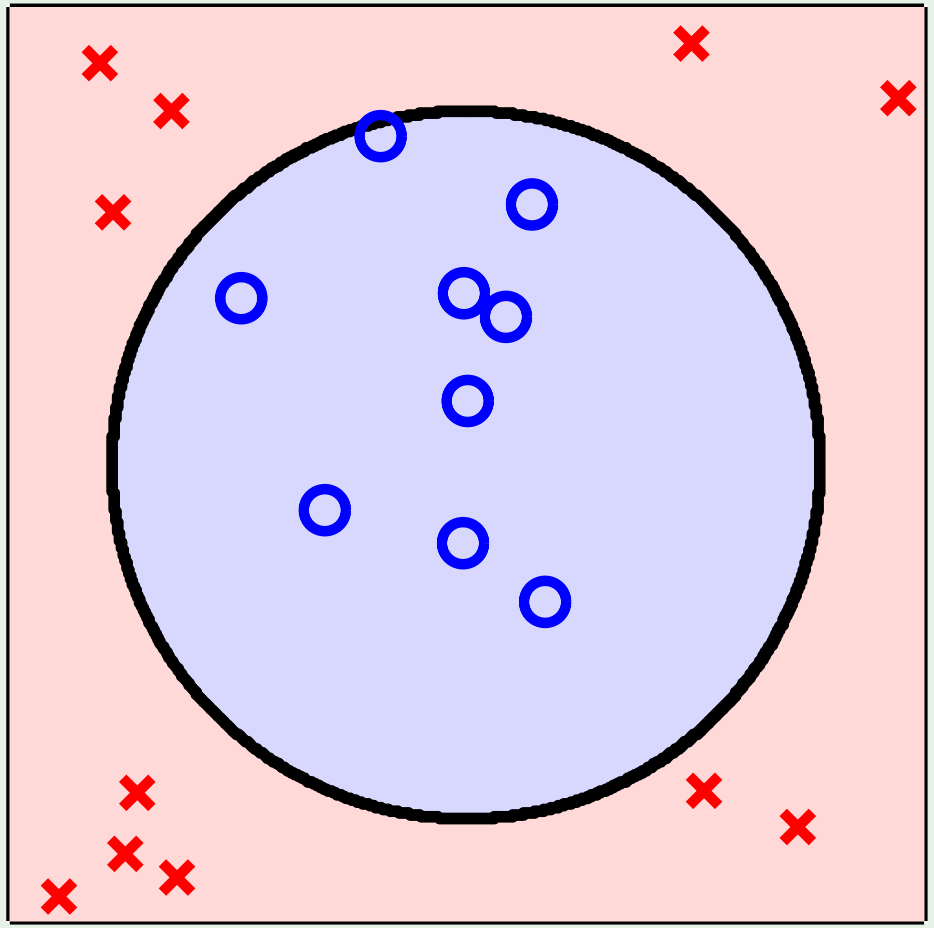
Second case

$$\mathbf{z} = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$$

Why not: $\mathbf{z} = (1, x_1^2, x_2^2)$

or better yet: $\mathbf{z} = (1, x_1^2 + x_2^2)$

or even: $\mathbf{z} = (x_1^2 + x_2^2 - 0.6)$



Lesson learned

Looking at the data *before* choosing the model can be hazardous to your E_{out}

Data snooping