Nonlinear transforms

\[ x = (x_0, x_1, \cdots, x_d) \xrightarrow{\Phi} z = (z_0, z_1, \cdots, z_d) \]

Each \( z_i = \phi_i(x) \) \quad \quad \quad z = \Phi(x)

Example: \( z = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2) \)

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

\[ \text{sign} \left( \tilde{w}^T \Phi(x) \right) \quad \text{or} \quad \tilde{w}^T \Phi(x) \]
The price we pay

\[ \mathbf{x} = (x_0, x_1, \ldots, x_d) \xrightarrow{\Phi} \mathbf{z} = (z_0, z_1, \ldots, z_{\tilde{d}}) \]

\[ \downarrow \quad \downarrow \]

\[ \mathbf{w} \quad \tilde{\mathbf{w}} \]

\[ d_{\text{VC}} = d + 1 \quad d_{\text{VC}} \leq \tilde{d} + 1 \]
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 $E_{\text{out}}$

Data snooping