Guided Tour of Machine Learning in Finance

The No Free Lunch Theorem

Igor Halperin

NYU Tandon School of Engineering, 2017

The Bias-Variance Tradeoff

$$\mathbb{E}\left[\left(y-\hat{f}(\boldsymbol{x})\right)^2\right]=(bias)^2+variance+noise$$

- **Bias-variance tradeoff:**
  - <u>Complex models</u> (more features) tend to have <u>low bias</u> and <u>high variance</u>
  - <u>Simple models</u> (less features) tend to have <u>high bias</u> and <u>low variance</u>
- Important to control **model complexity** to have an optimal tradeoff!

The No Free Lunch Theorem

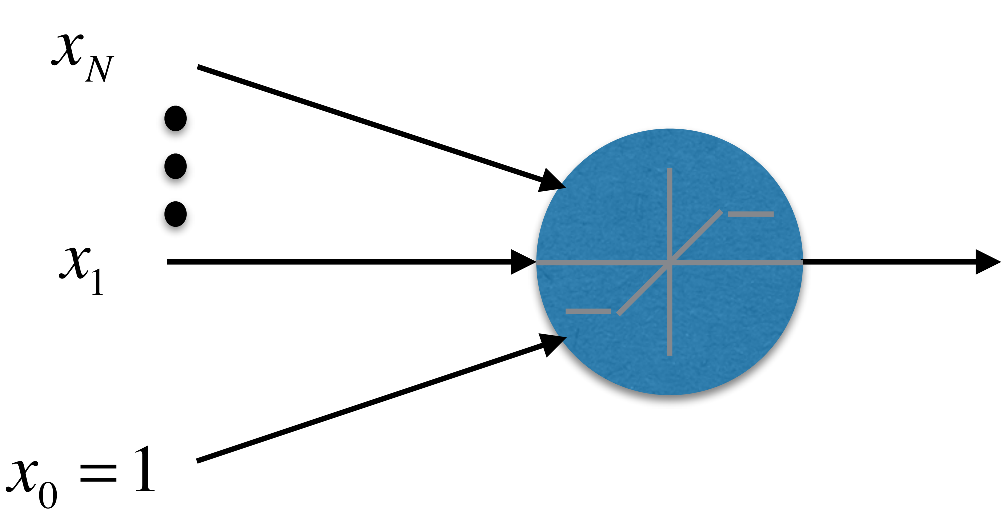
- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.

The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.

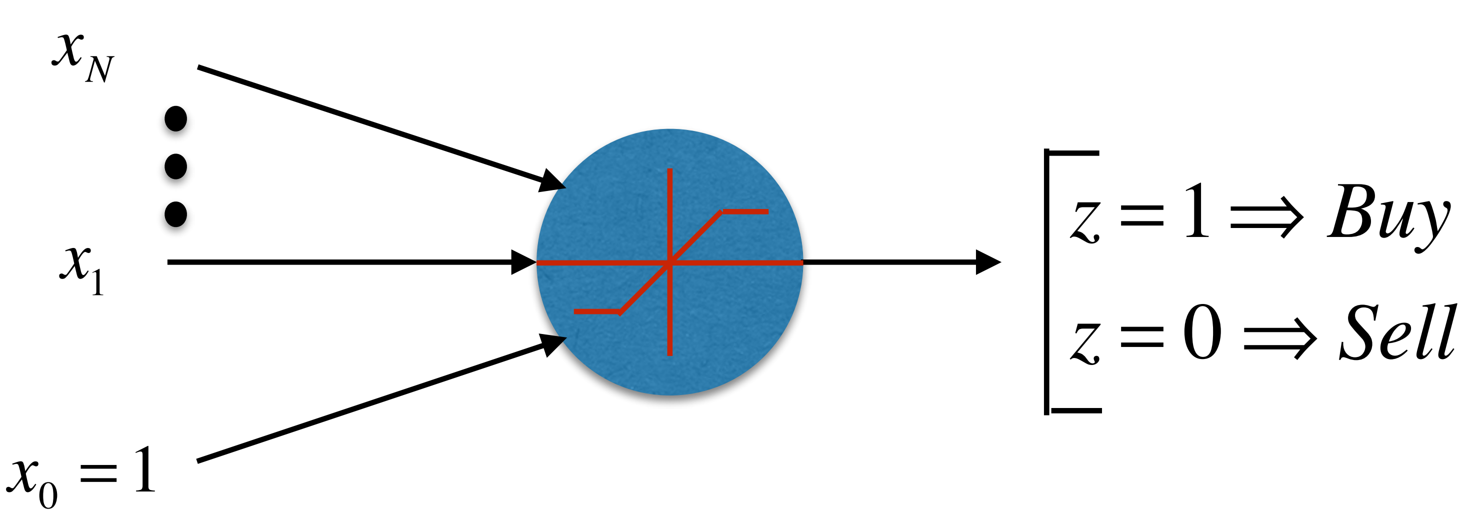
The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.
- **Example**: Hot stock or not hot stock?



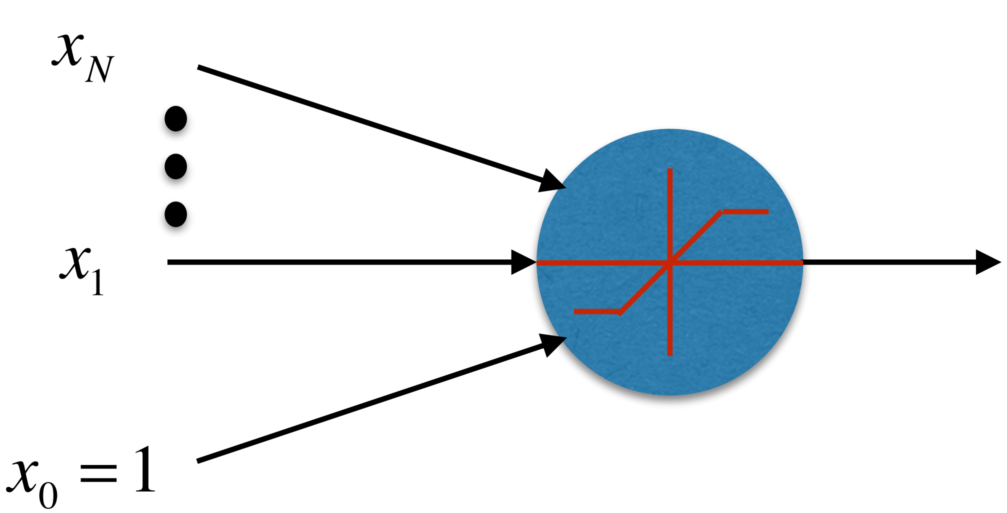
The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.
- **Example**: Hot stock or not hot stock?



The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.
- **Example**: Hot stock or not hot stock?



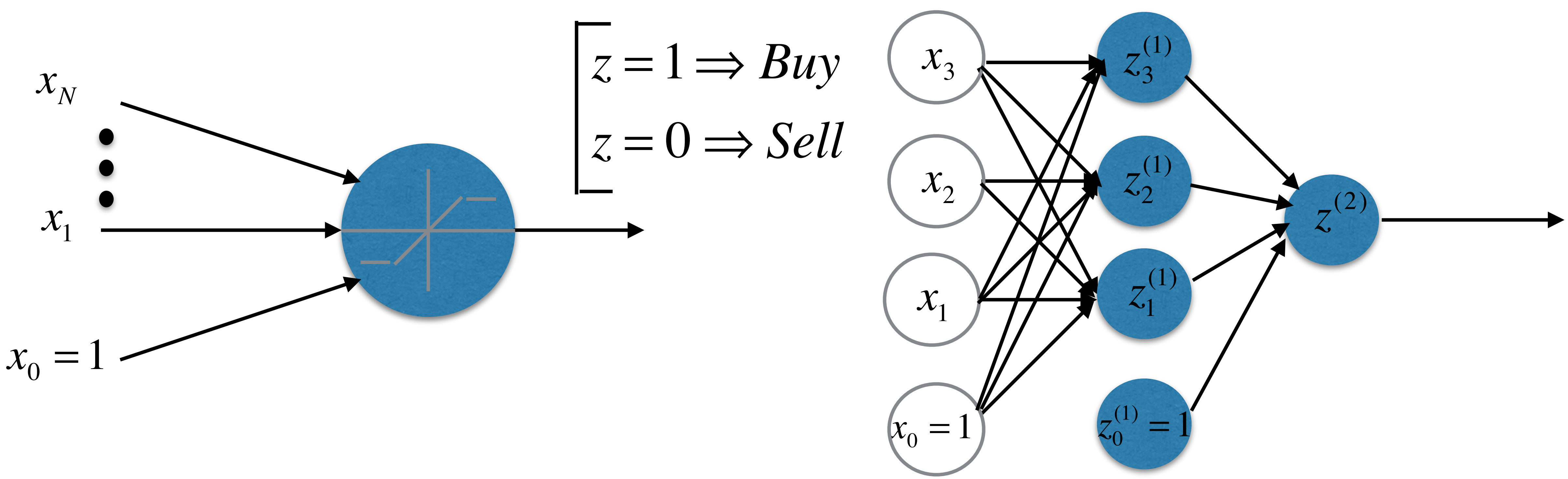
$$\begin{cases} z = 1 \Rightarrow Buy \\ z = 0 \Rightarrow Sell \end{cases}$$

Adjustable parameters

$$E.g.: \sigma(z = \sum_{i=0}^{N} \theta_n x_n) = 1/(1+\exp(-z))$$

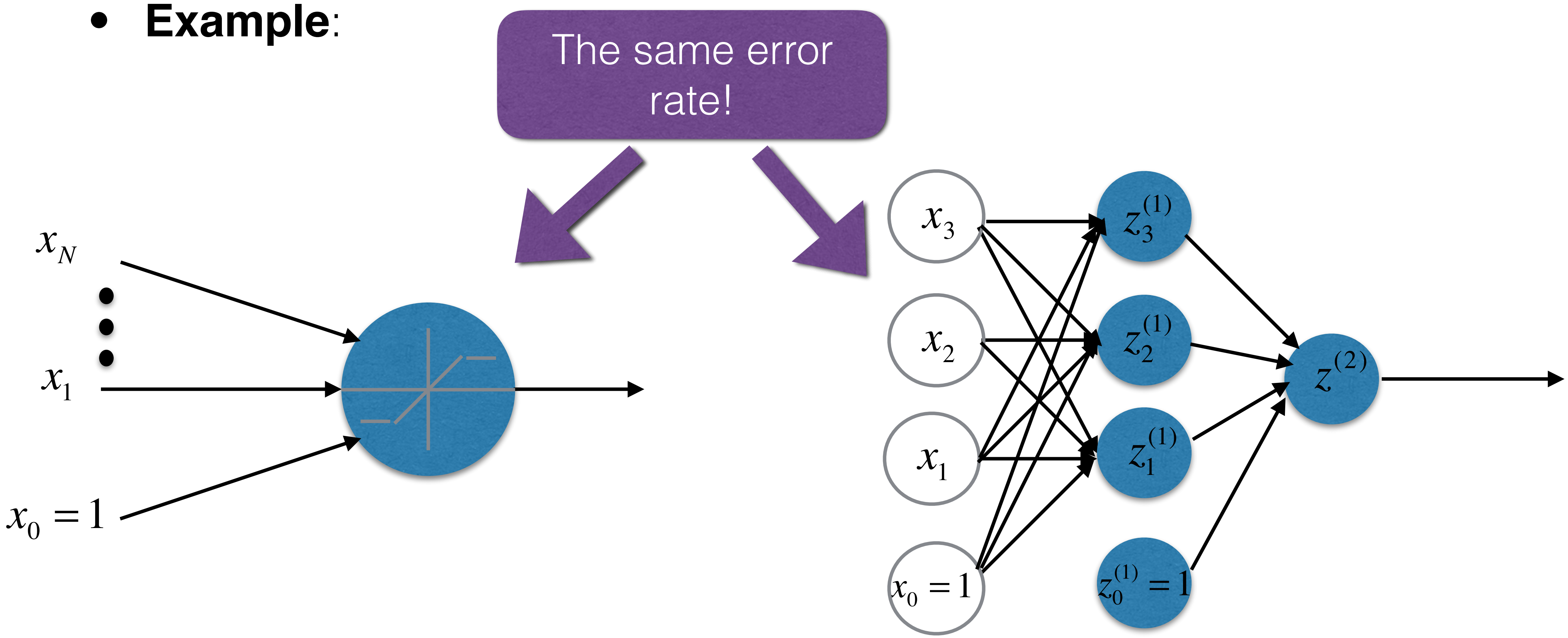
The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.
- **Example**:



The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: no single ML classification algorithm can be universally better than any other one on all domains.
- **Example**:



The No Free Lunch Theorem

- The **No Free Lunch Theorem** (Wolpert, 1996): all classifiers have the **same error rate** when averaged over **all** possible data-generating distributions.
- Meaning: <u>no single ML classification algorithm can be universally better than any other one on all domains</u>.
- **A** p**ractical remedies**:
  - ✦ Look for ML algorithms that perform best on certain <u>classes of domains</u> (characterized by a data-generating distribution p_{data}).