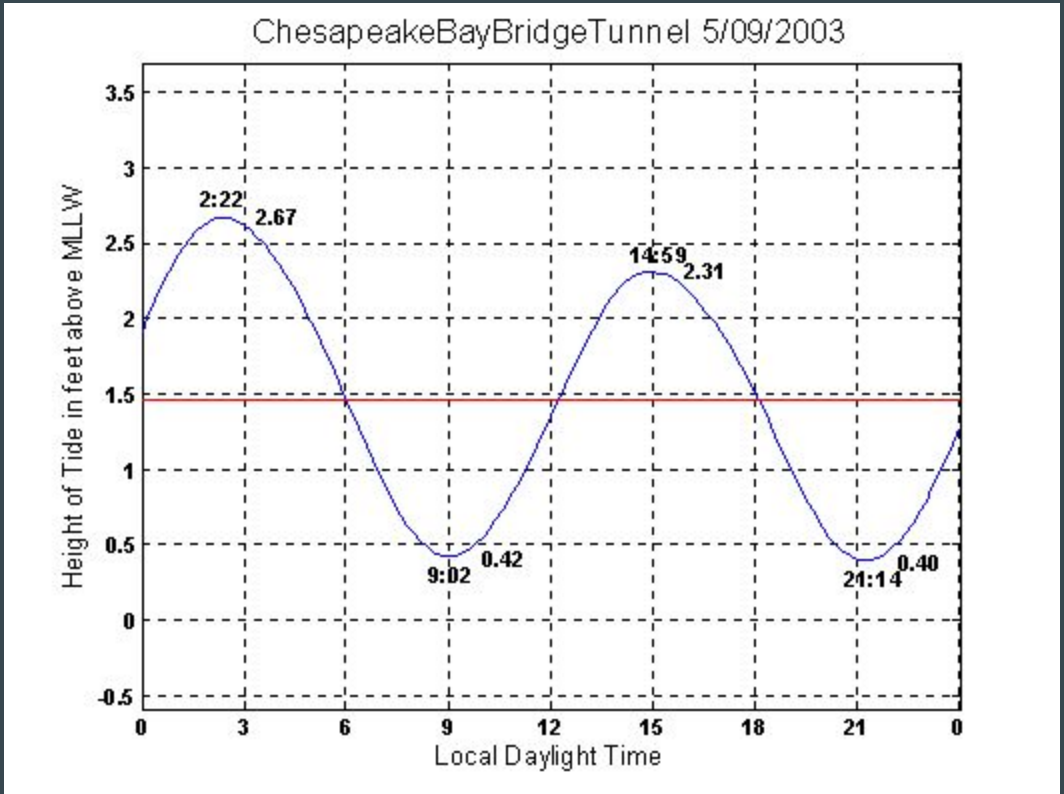
Seminário: Viés e Variância

José Geraldo Fernandes

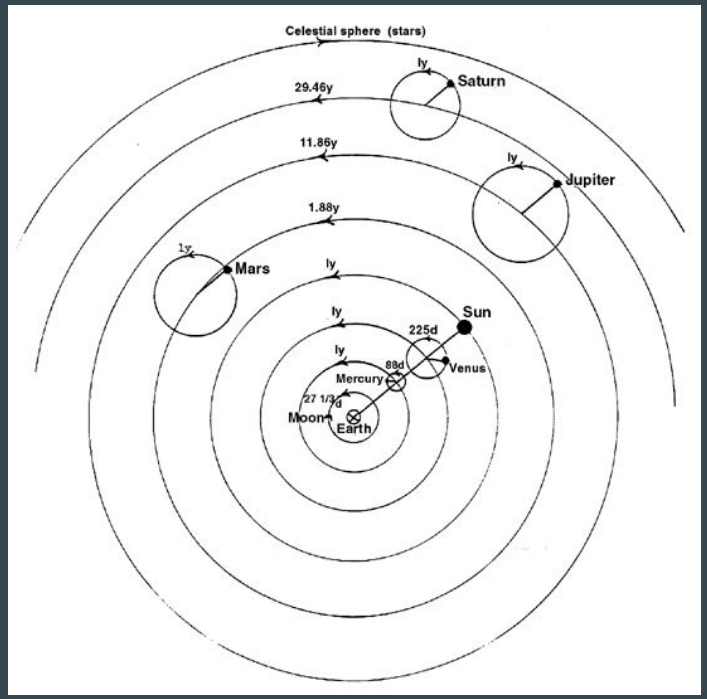
Geman, Stuart, Elie Bienenstock, and René Doursat. "Neural networks and the bias/variance dilemma." Neural computation 4.1 (1992): 1-58.

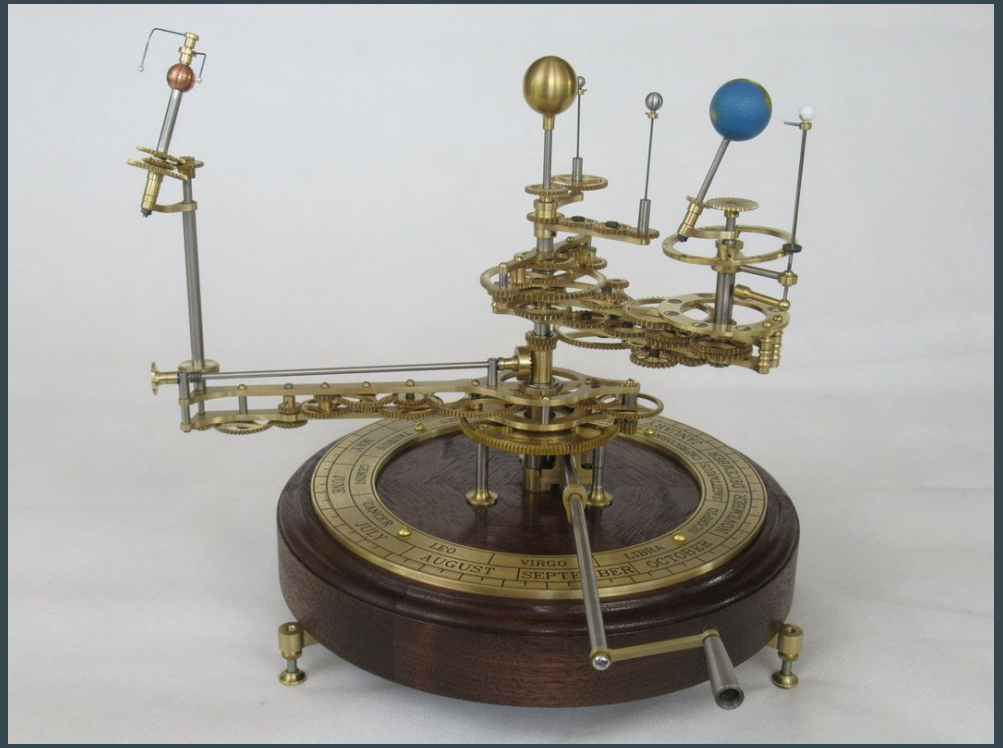
Capacidade dos Modelos

Maré


ChesapeakeBayBridgeTunnel 5/09/2003
Height of Tide in feet above MLLW
Local Daylight Time
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
0
3
6
9
12
15
18
21
0
2:22
2.67
9:02
0.42
14:59
2.31
21:14
0.40

Astronomia


Celestial sphere (stars)
1y
Saturn
29.46y
11.86y
1y
Jupiter
1.88y
1y
Mars
1y
1y
Sun
225d
1y
Venus
88d
Mercury
27 1/3d
Moon
Earth


LEO
VIRGO
LIBRA
AUGUST
OCTOBER

Modelos Não-Paramétricos

Tabula Rasa

- Fronteira arbitrária, aproximador universal
- Poupa o estudo de conhecimento *a priori*
- “Let the data speak for itself”

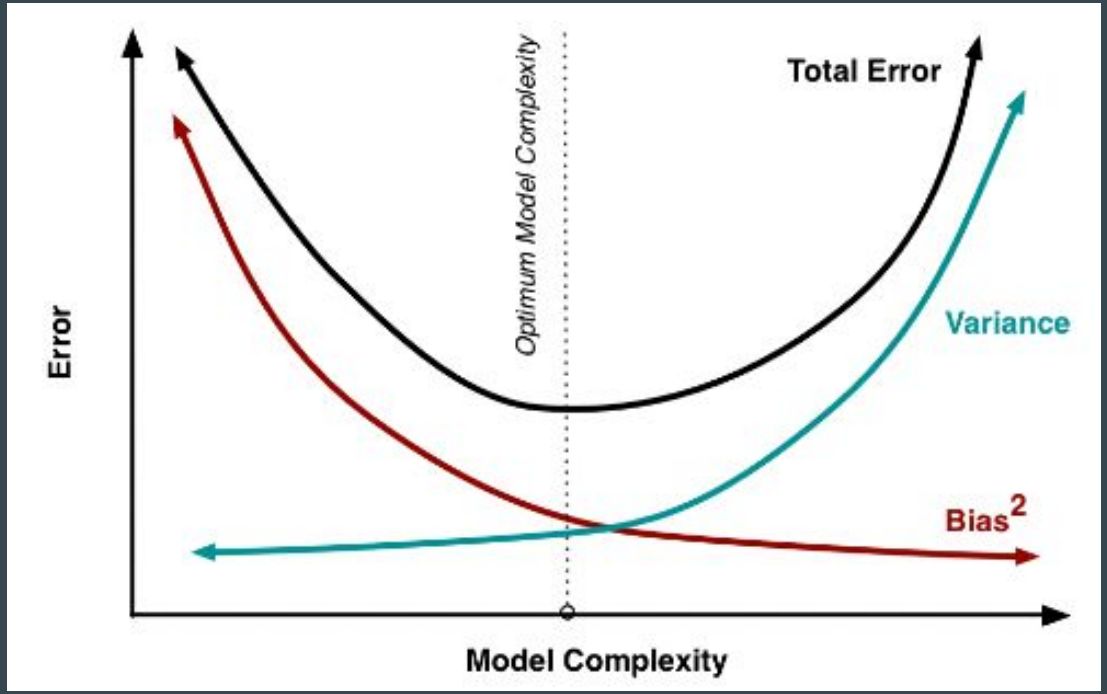
Modelo “Geral”

$$y = h(\mathbf{x}) + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$
\begin{aligned}
&E_{\mathcal{D}}\left[(f(\mathbf{x};\mathcal{D}) - E[y \mid \mathbf{x}])^2\right] \\
&\quad = E_{\mathcal{D}}\left[((f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right]) + (E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}]))^2\right] \\
&\quad = E_{\mathcal{D}}\left[(f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right])^2\right] + E_{\mathcal{D}}\left[(E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}])^2\right] \\
&\qquad + 2E_{\mathcal{D}}\left[(f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right])\,(E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}])\right] \\
&\quad = E_{\mathcal{D}}\left[(f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right])^2\right] + (E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}])^2 \\
&\qquad + 2E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right]\right] \cdot (E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}]) \\
&\quad = (E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right] - E[y \mid \mathbf{x}])^2 \qquad \text{“bias”} \\
&\qquad + E_{\mathcal{D}}\left[(f(\mathbf{x};\mathcal{D}) - E_{\mathcal{D}}\left[f(\mathbf{x};\mathcal{D})\right])^2\right] \qquad \text{“variance”}
\end{aligned}
$$

Bias-Variance Tradeoff

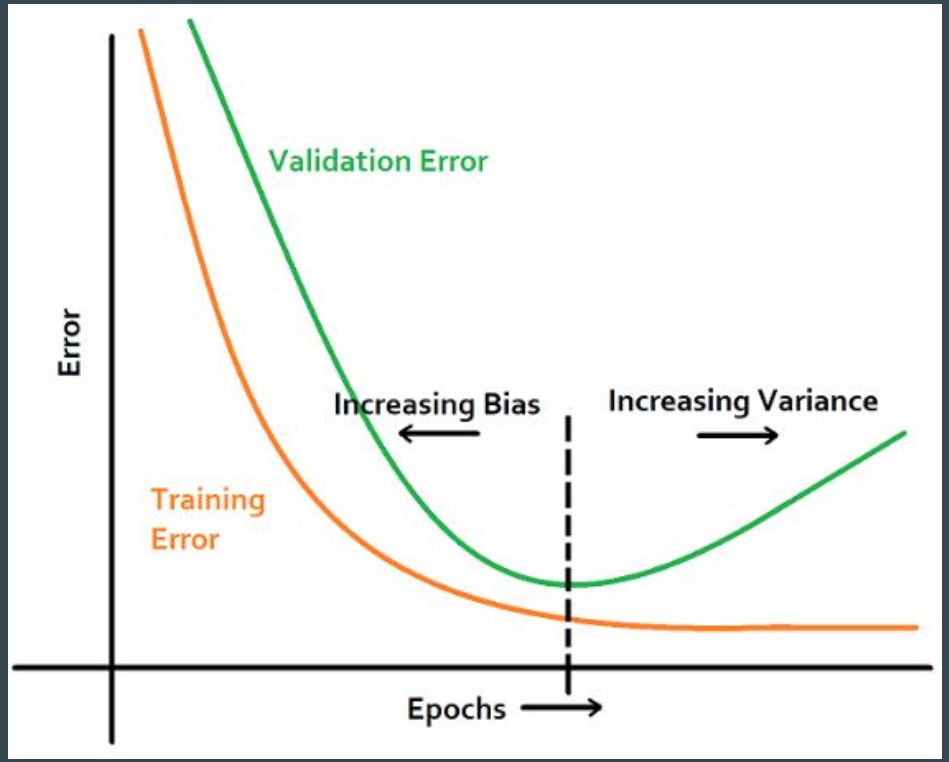


Otimização Multiobjetivo

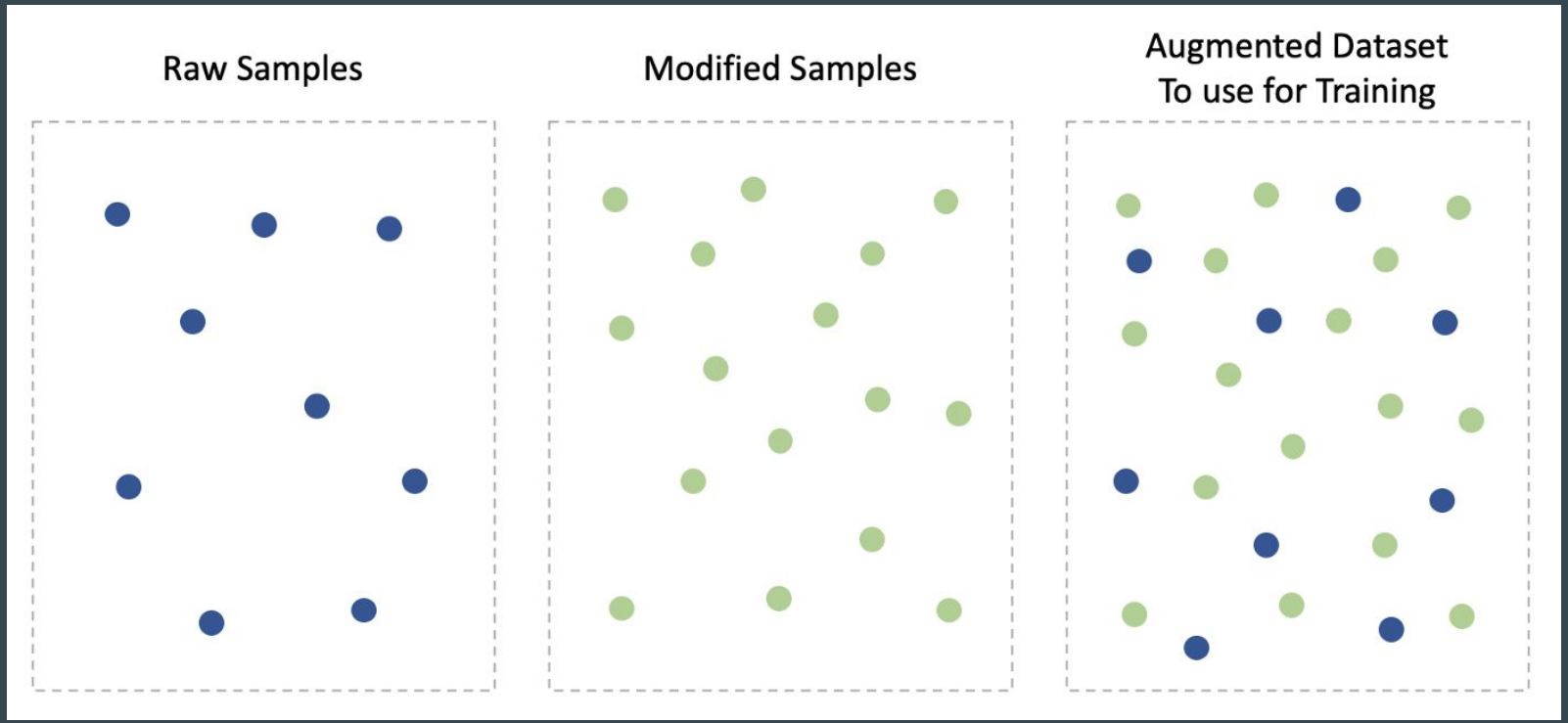
Regularização

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

Cross Validation / Early Stopping

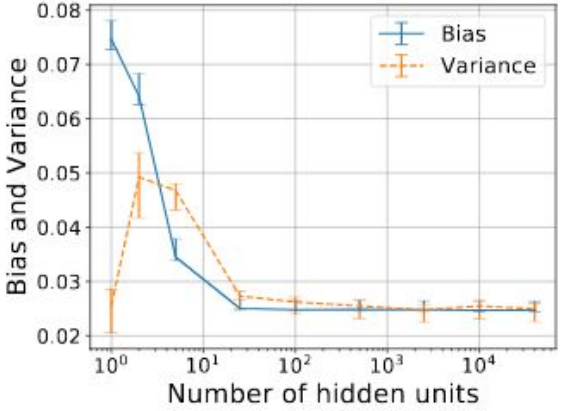


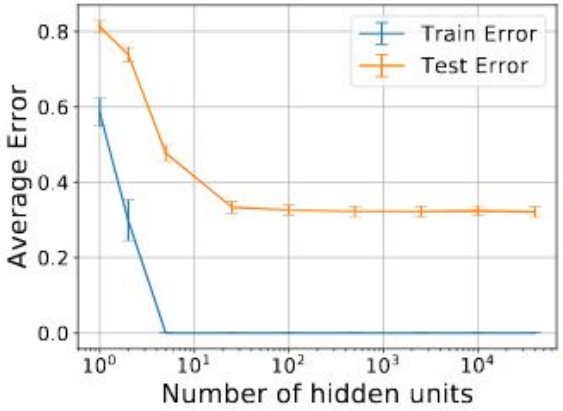
Data Augmentation

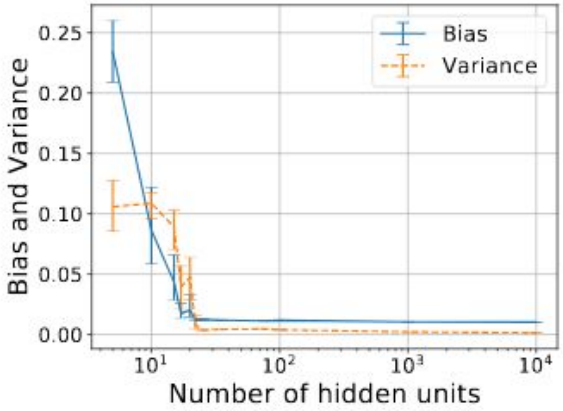


Uma Visão “Moderna”

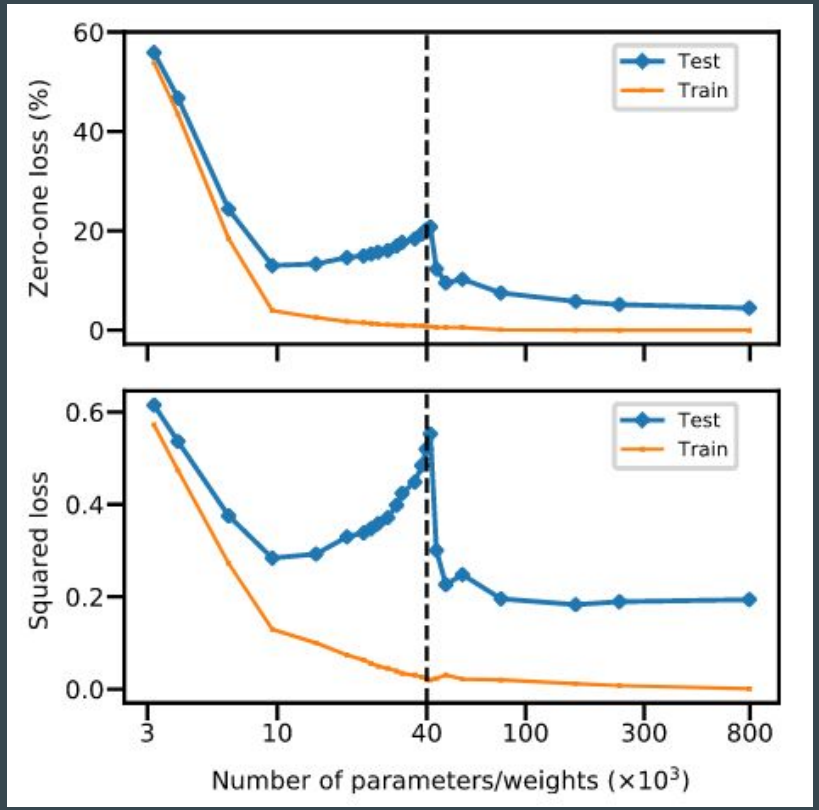
Comportamento Estranho



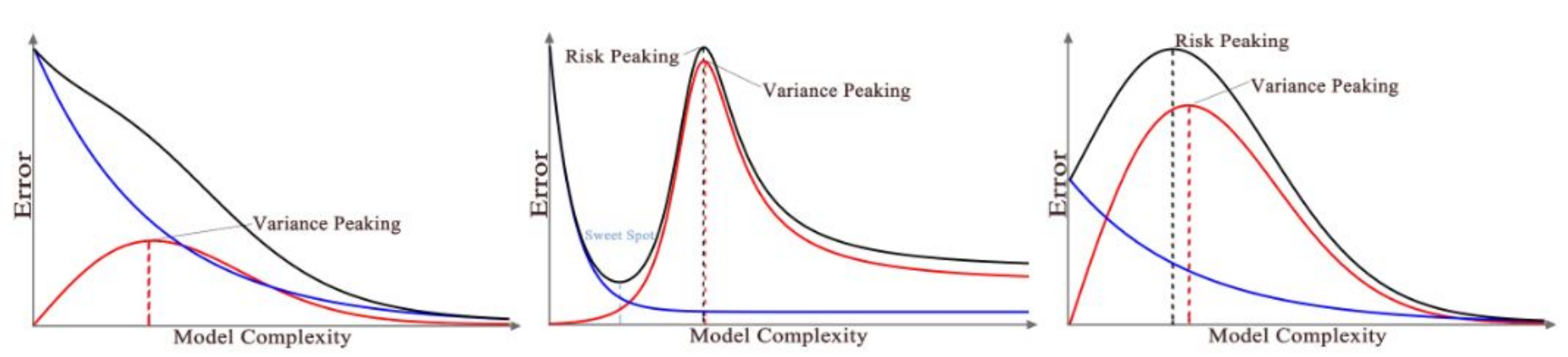




Comportamento Estranho



Comportamento Estranho



Discussão

Discussão

- Comportamento não previsto da variância
- Interpolação
- Quantidade de dados e cobertura
- Como medir a complexidade/capacidade dos modelos
  - Número de neurônios
  - Número de camadas
- Como medir o viés e variância
  - Validação cruzada
  - Medição direta
- Regularização implícita

Obrigado