



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

符号回归与深度学习分析数据

汇报人：李庆梦 博士后

合作导师：赵丽娜 研究员

部门/课题组：多学科研究中心/AI分子组

2024.10.16

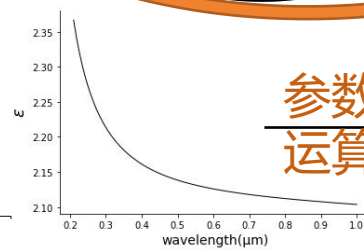
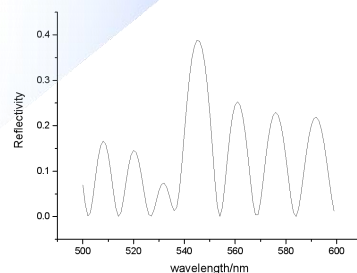
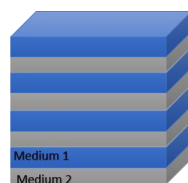
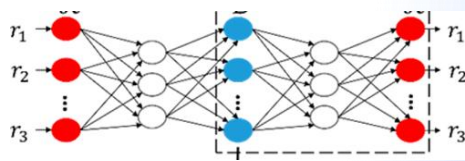
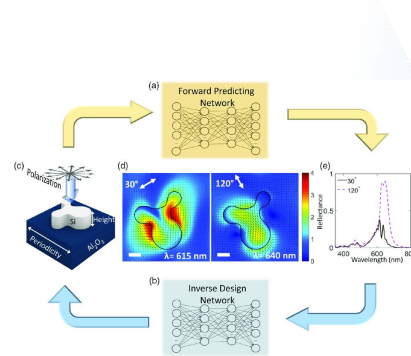


1.1.符号回归的功能与优势--表征光学折射率相关函数

i) 结构设计^[1]

ii) 表征等离子体谐振器

“黑箱”



$\epsilon = ?$

机器学习

深度学习
神经网络

符号回归

元启发式优化算法
进化算法 粒子群优化

预定模型^[2]

$$\epsilon(\omega) = 1 + \sum_{j=1}^k \frac{f_j \omega_p^2}{(\omega_j^2 - \omega^2) + i\omega\Gamma_j}$$

拟合参数

$$\epsilon(\omega | \omega_j, f_j, \Gamma_j)$$

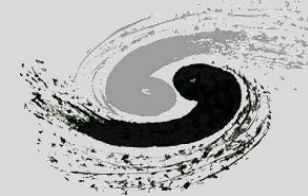
参数
运算符

可视化模型:

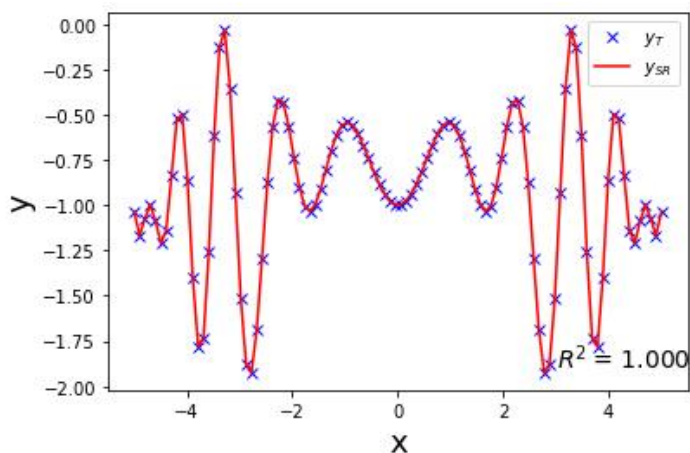
$$\epsilon(\omega) = 4.3 + \frac{0.3\omega + i(0.5\omega - 0.9)}{\omega^2 - i(5.1\omega - 0.3)}$$

[1] I. H. Malitson, "Interspecimen Comparison of the Refractive Index of Fused Silica*,†," J. Opt. Soc. Am. **55**, 1205-1209 (1965)

[2] A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski. Optical properties of metallic films for vertical-cavity optoelectronic devices, *Appl. Opt.* **37**, 5271-5283 (1998)

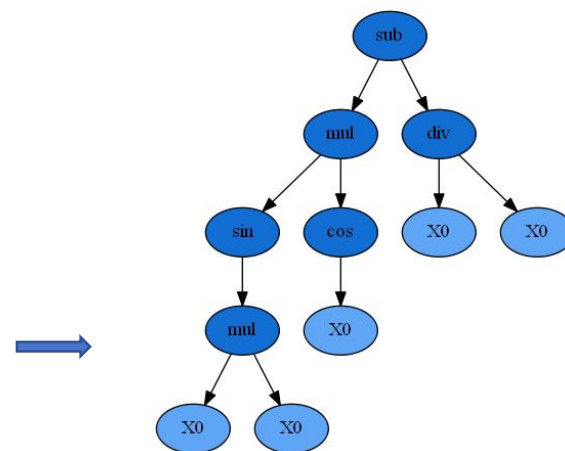
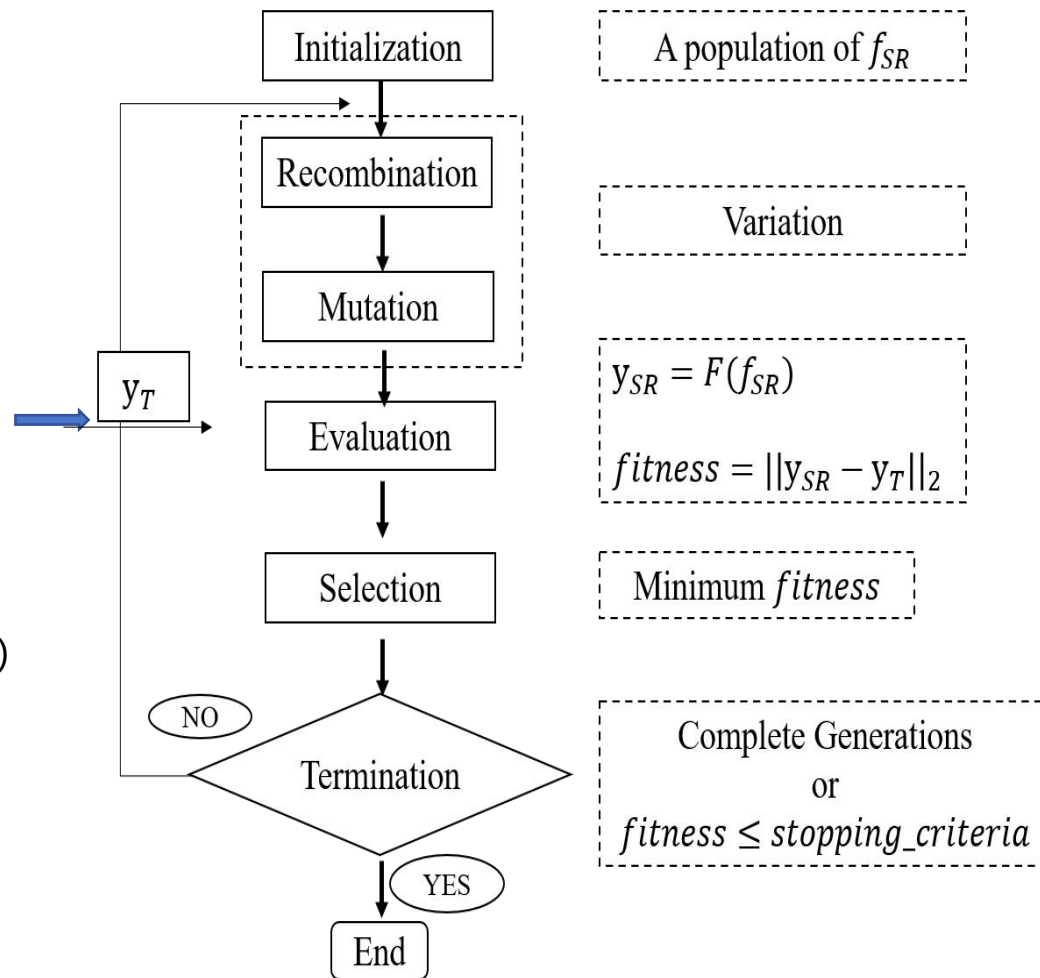


1.2. 基于遗传编程的符号回归方法原理



(a) 输入数据 (蓝x) 和输出函数数据 (红线)

目标函数 (蓝色x) :
 $y_T = \sin(x^2) \cos(x) - 1$



(b) 语法树

输出函数:

$$y_{SR} = \sin(X0^2) \cos(X0) - \frac{X0}{X0}$$

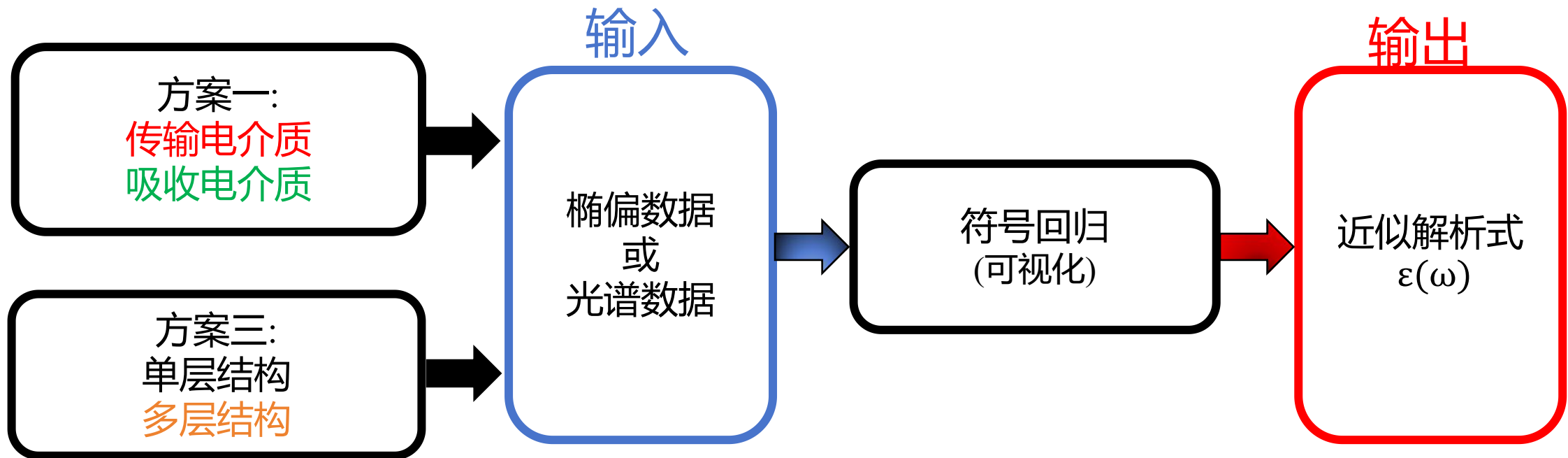
Sympy

简化函数:

$$y_{SR} = \sin(X0^2) \cos(X0) - 1$$

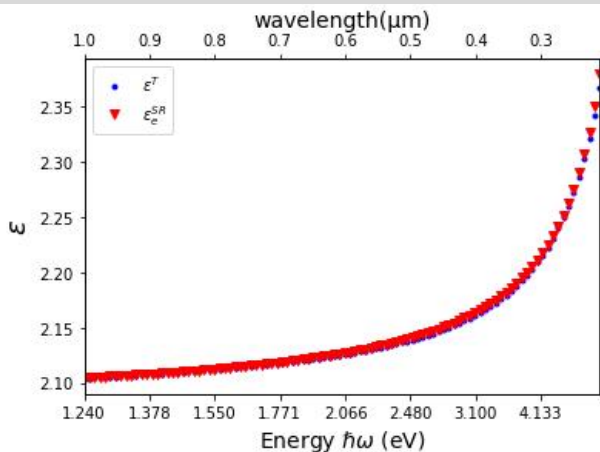


1.3.案例：使用符号回归表征光学折射率相关函数





1.4.探索直接关系解析式



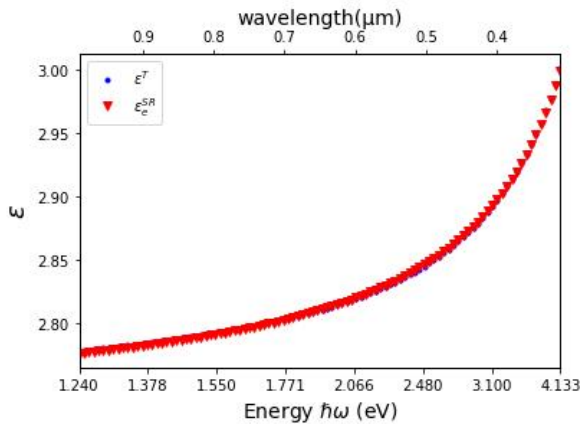
(a) 传输电介质SiO₂ [1]函数

目标函数（蓝点输入数据）：

$$\epsilon^T = 1 + \frac{0.6961663\lambda^2}{\lambda^2 - 0.0684043^2} + \frac{0.4079426\lambda^2}{\lambda^2 - 0.1162414^2} + \frac{0.8974794\lambda^2}{\lambda^2 - 9.896161^2}$$

输出函数:

$$\epsilon^{SR} = 2.078 + \frac{0.023}{\lambda - 0.134}$$



(b) 传输电介质Al₂O₃ [2]函数

无目标函数:

输出函数：

$$\epsilon^{SR} = 2.727 + \frac{0.041}{\lambda - 0.16}$$

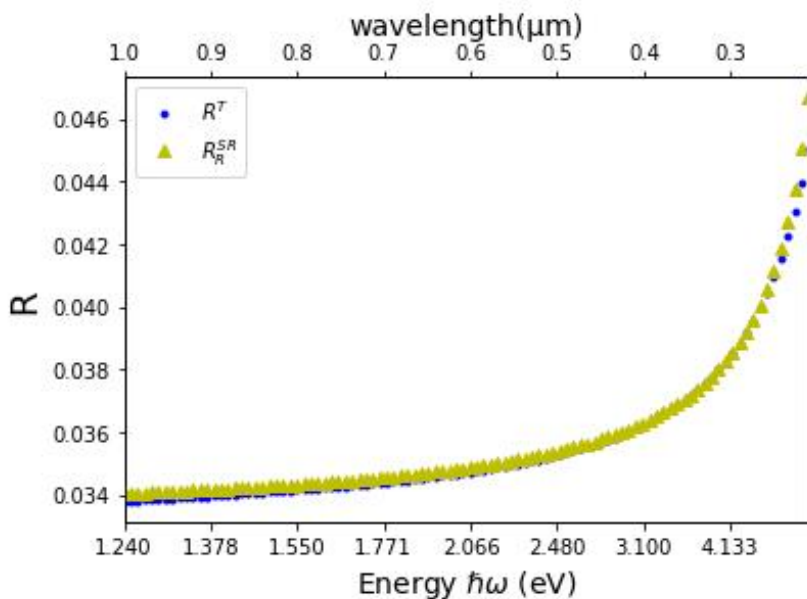
[1] I. H. Malitson, "Interspecimen Comparison of the Refractive Index of Fused Silica*,†," J. Opt. Soc. Am. **55**, 1205-1209 (1965)

[2] I. H. Malitson and M. J. Dodge. Refractive Index and Birefringence of Synthetic Sapphire, [J. Opt. Soc. Am. **62**, 1405 \(1972\)](#)

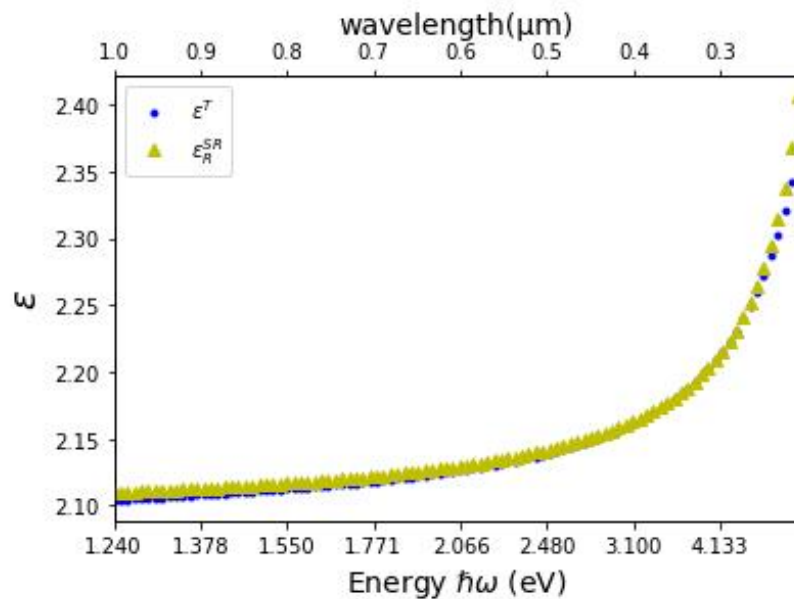


1.5.探索间接关系解析式-嵌入界面结构模型

➤ 基于反射光谱来表征传输电介质SiO₂数据的解析式



(a)反射光谱



(b)介电函数.

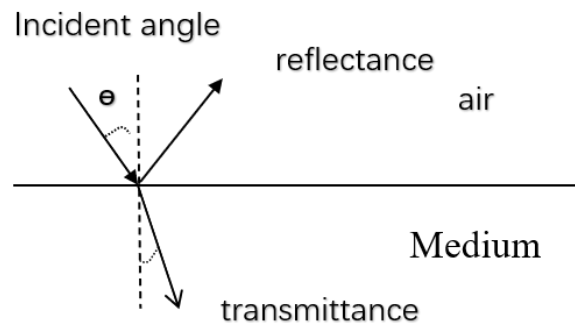


图 (a) 中反射光谱 R^T

符号回归
"+, -, *, /"

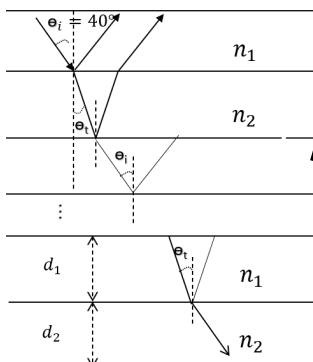
输出函数:

$$\epsilon^{SR} = 2.093 + \frac{0.012}{\lambda^2}$$



1.6.探索间接关系解析式-嵌入周期性多层纳米结构模型

➤ 基于反射光谱模拟CaCO₃电介质函数

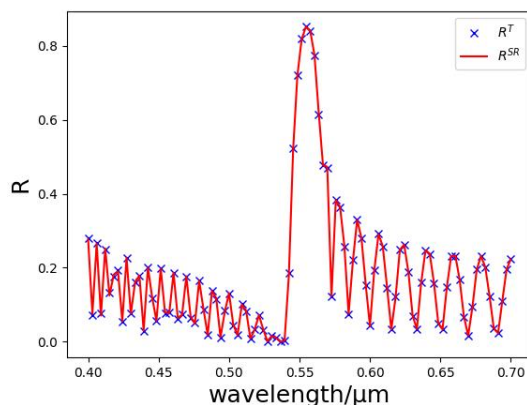


Berreman 4x4 矩阵法

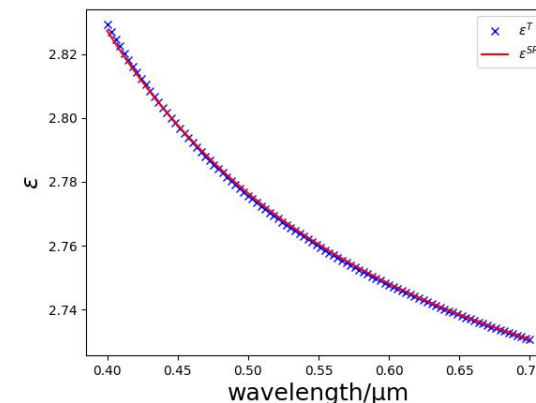
符号回归

输出

输入



(a) 反射光谱



(b) 介电函数

输出函数:

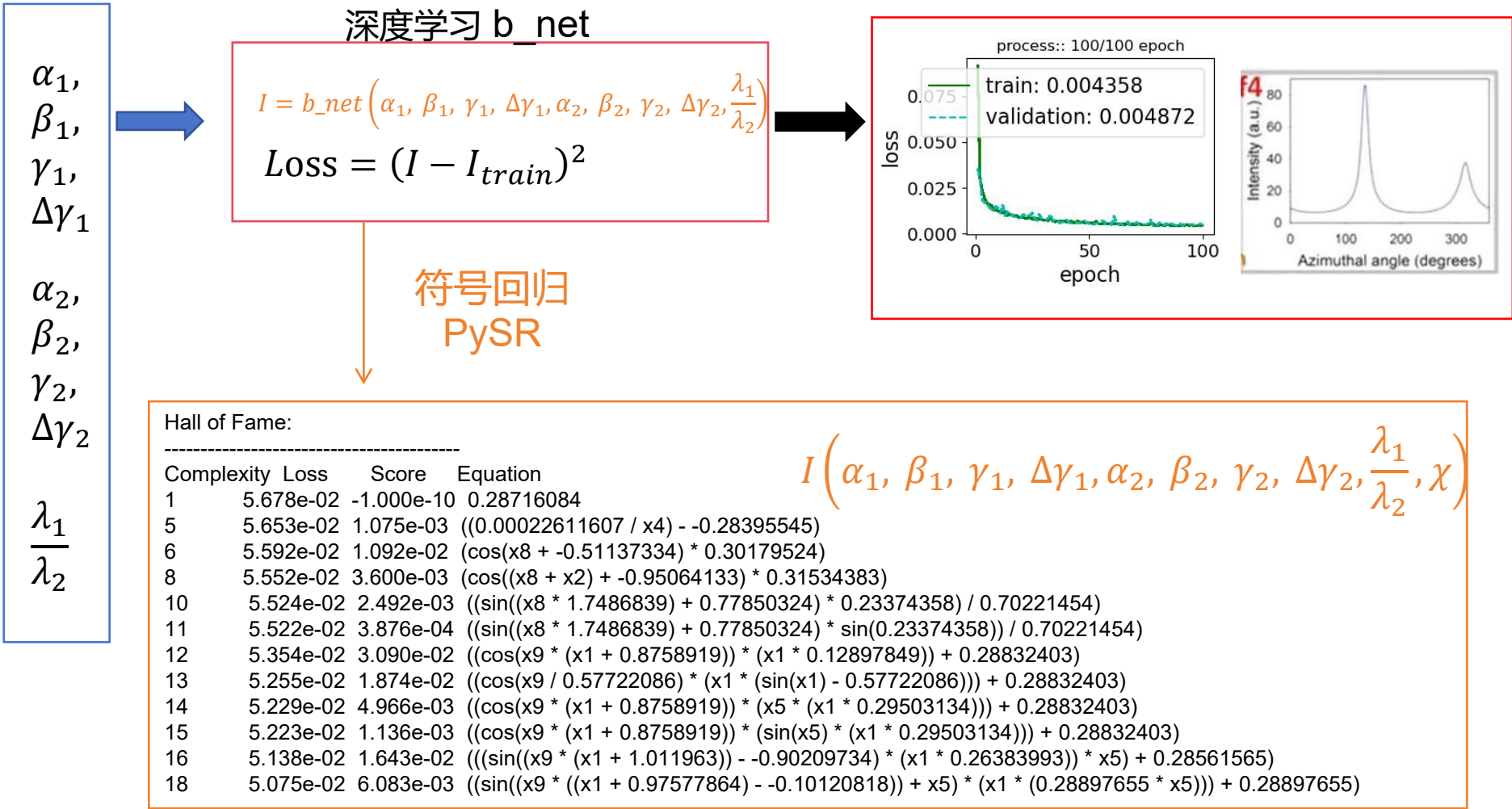
$$\epsilon^{SR} = 2.714 + \frac{0.011\lambda + 0.02}{\lambda^2}$$

目标函数:

$$\epsilon^T = 1.73358749 + \frac{0.96464345\lambda^2}{\lambda^2 - 0.0194325203} + \frac{1.82831454\lambda^2}{\lambda^2 - 120}$$



2.1.符号回归与深度学习分析数据--探索数据解析式



- I. 深度学习训练由纤维取向（9个参数）到衍射强度曲线的预测模型b_net
- II. 用符号回归将b_net用解析式表达

[1] Zhang Y, Paris O, Terrill N J, et al. Uncovering three-dimensional gradients in fibrillar orientation in an impact-resistant biological armour[J]. Scientific reports, 2016, 6(1): 1-13.

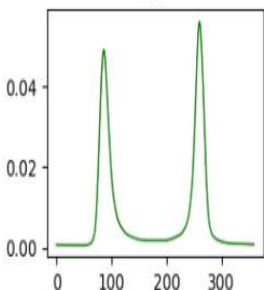
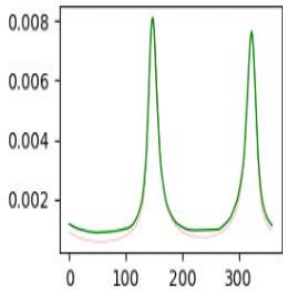
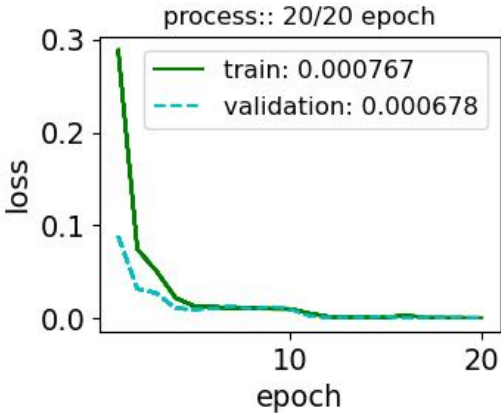


2.2.符号回归与深度学习分析数据--降维数据后探索解析式

$\alpha_1,$
 $\beta_1,$
 $\gamma_1,$
 $\Delta\gamma_1,$
 λ_1
 χ
 $\alpha_2,$
 $\beta_2,$
 $\gamma_2,$
 $\Delta\gamma_2,$
 λ_2

深度学习 sub_net

$$I_1 = \text{sub_net}(\alpha_1, \beta_1, \gamma_1, \Delta\gamma_1, \lambda_1, \chi)$$
$$I_2 = \text{sub_net}(\alpha_2, \beta_2, \gamma_2, \Delta\gamma_2, \lambda_2, \chi)$$
$$\text{Loss} = (I_1 + I_2 - I_{\text{train}})^2$$



Hall of Fame:

Complexity	Loss	Score	Equation
1	9.309e-05	2.344e-07	0.008001122
4	8.874e-05	1.594e-02	(0.005703777 / cos(x4))
5	8.841e-05	3.790e-03	(0.005703777 - (x4 * -0.0044588447))
6	8.821e-05	2.213e-03	(sin(x4 + 0.5840846) * 0.010172627)
7	8.473e-05	4.029e-02	((0.6054646 + (x4 * x2)) * 0.010172627)
8	8.295e-05	2.125e-02	(cos((x2 - 0.4603302) * 2.850666) / 82.97504)
10	7.934e-05	2.220e-02	(cos(x2 - x4) / (74.98859 + (1.007408 / x4)))
11	7.933e-05	2.158e-04	(cos(x2 - x4) / (74.98859 + (cos(x4) / x4)))
12	7.480e-05	5.883e-02	(cos((x2 - 0.49039188) * 2.850666) / ((1.1763369 / x4) + 74.98859))
14	7.429e-05	3.369e-03	(cos((x2 - 0.4819017) * -3.0198097) / ((74.98859 + (1.0184219 / x4)) / 1.1982089))
15	7.382e-05	6.377e-03	(cos((x2 - 0.4819017) * -3.0198097) / ((74.98859 * cos(-0.5319049)) + (1.0593079 / x4)))
16	7.371e-05	1.518e-03	(cos((x2 - 0.49039188) * 2.850666) / ((74.98859 * cos(-0.51001185)) + (1.3720603 / sin(x4))))
17	7.326e-05	6.155e-03	(cos((x2 - 0.49039188) * 2.850666) / ((74.98859 * cos(x1 / 1.4016266)) + (1.3720603 / x4)))
19	7.264e-05	4.243e-03	(cos((x2 - 0.49039188) * 2.850666) / ((cos(x1 / 1.4016266) * 74.98859) + ((0.60186714 + x3) / x4)))
20	7.257e-05	9.369e-04	(cos((x2 - 0.49039188) * 2.850666) / ((cos(x1 / 1.4016266) * 74.98859) + ((cos(-0.6559321) + x3) / x4)))

SR

$$I_i(\alpha_i, \beta_i, \gamma_i, \Delta\gamma_i, \lambda_i, \chi)$$
$$I = I_1 + I_2$$

- I. 训练低维子预测模型sub_net
- II. 用符号回归将sub_net用解析式表达

[1] Zhang Y, Paris O, Terrill N J, et al. Uncovering three-dimensional gradients in fibrillar orientation in an impact-resistant biological armour[J]. Scientific reports, 2016, 6(1): 1-13.



➤ 总结：

- I. 深度学习探索高维数据的“黑箱”规律
- II. 神经网络实现高维数据降维
- III. 符号回归探索低维数据“玻璃盒”解析关系式

➤ 展望：

- 结合符号回归与深度学习探索更多领域的数据解析式规律



中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

谢谢批评与指正!