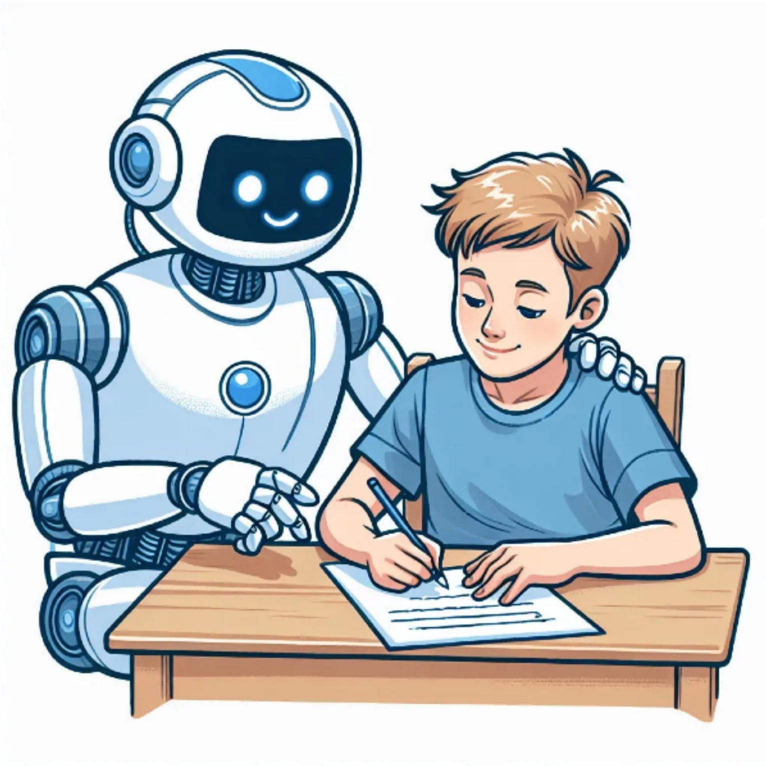
AI & Machine Learning in Physics



PHYS3151 (6 credits)

Time & Place : Tue 16:30-17:20; 17:30-18:20
Fri 17:30-18:20

Teachers: Zi Yang Meng (zymeng@hku.hk), HOC 231

Tutor: Tim-Lok Chau (Justin) (justintlchau@connect.hku.hk)

AI & Machine Learning in Physics

Teaching Materials:

https://quantummc.xyz/teaching/hku-phys3151-machine-learning-in-physics-2025/

Slides / Reading materials

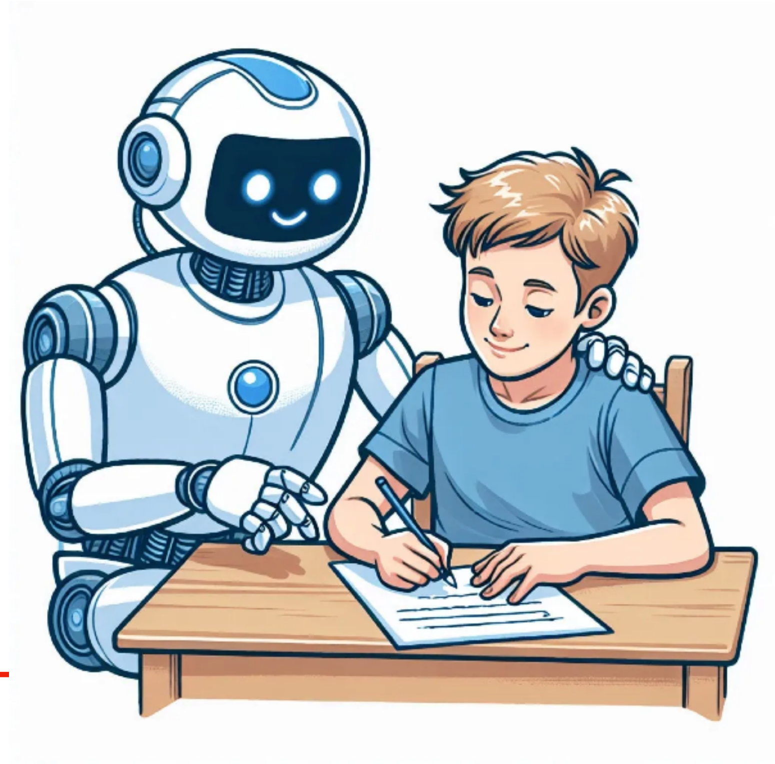
Python notebooks

Assignments

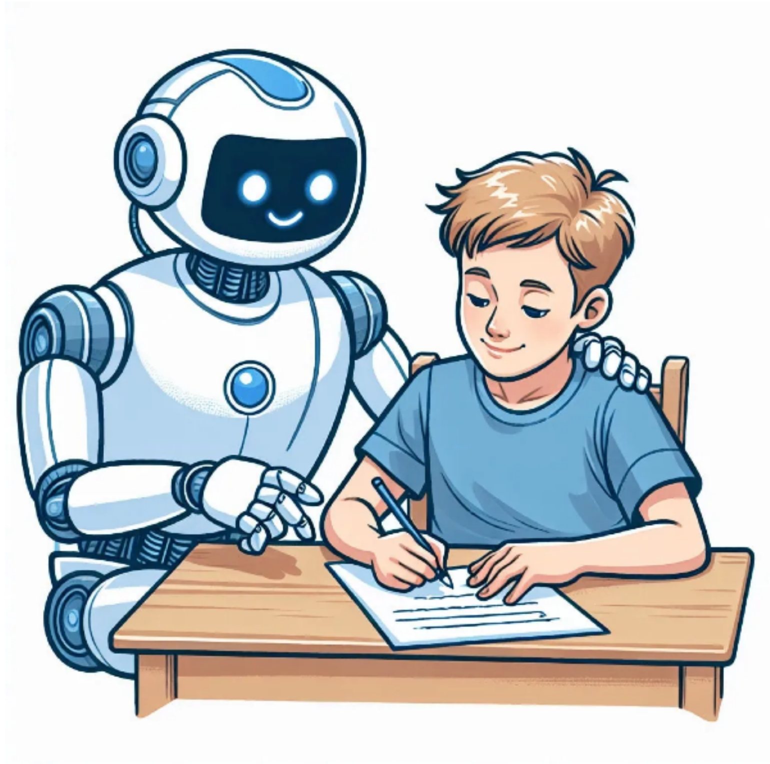
Assessment Methods and Weighting

- Assignments   30%
- Presentation   20%
- Project report 20%
- Exam.   30%

Content



Content



Multivariate Linear Regression

- Ethem Alpaydin, Introduction to Machine Learning, Third Edition, MIT Press 2014

Chap. 4. Parametric Methods
    4.6 Regression
    4.7 Bias/Variance Dilemma
Chap. 5. Multivariate Methods
    5.1 Multivariate data
    5.8 Multivariate Regression

- Linear algebra books:

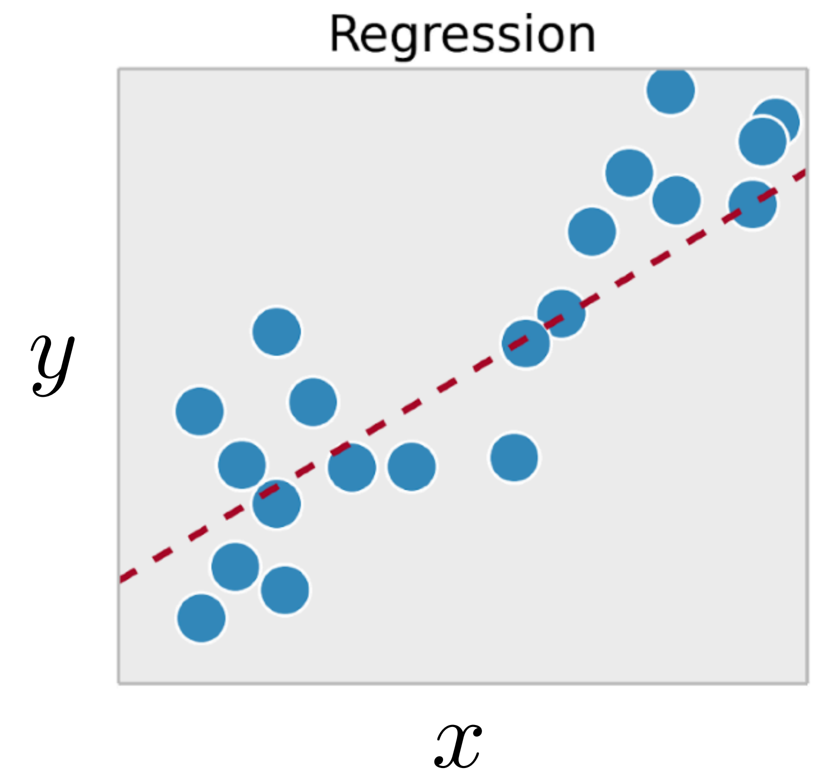
Gilbert Strang, Linear Algebra and its Applications, 1988
David Harville, Matrix Algebra from a Statistician's Perspective, 1997, Springer

- Most important

https://quantummc.xyz/teaching/hku-phys3151-machine-learning-in-physics-2025/

Multivariate Linear Regression

Regression: $y = h_\Theta(x) = \Theta \cdot x$



Statistics	Machine Learning	Notation	Remarks
independent variable	feature	$x_j^{(i)}$	$j = 1, \cdots, N$
dependent variable	outcome	$y^{(i)}$	
sample	example	$(x^{(i)}, y^{(i)})$	$i = 1, \cdots, M$
model	hypothesis	$h_\theta(x)$	
parameter	parameter	θ_j	
intercept	bias	θ_0	

$$\{(x_j^{(i)}, y^{(i)}), \theta_j\};\; j = 1, 2, \cdots, N;\; i = 1, 2, \cdots, M; N < M$$

$$y^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_N x_N^{(i)}$$

$$\mathbb{R}^{M\times(N+1)} \quad \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_N^{(2)} \\ & \vdots & & \vdots & \\ 1 & x_1^{(M)} & x_2^{(M)} & \cdots & x_N^{(M)} \end{bmatrix} \cdot \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(M)} \end{bmatrix}$$

$$\underline{\underline{X}} \cdot \underline{\Theta} = \underline{Y}$$

Multivariate Linear Regression

Basic assumption: examples are independent and identically distributed (i.i.d.).

Likehood function

$$\mathcal{L}(\Theta|X) = \sum_{i=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y^{(i)} - \theta^T \cdot x^{(i)})^2}{2\sigma^2})$$

$$\Theta^* = \arg\min_{\Theta}\{\sum_{i=1}^{M} \frac{(y^{(i)} - \theta^T \cdot x^{(i)})^2}{2M}\}$$

Cost / Loss function

$$J(\Theta) = \frac{1}{2M}\sum_{i=1}^{M}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Convex optimisation

$$\begin{aligned}\Delta_\theta J(\theta) &= \nabla_\theta \cdot \nabla_\theta J(\theta) \\ &= \nabla_\theta \cdot \nabla_\theta \left[\frac{1}{2M}\sum_{i=1}^{M}\left(\theta^T x^{(i)} - y^{(i)}\right)^2\right] \\ &= \nabla_\theta \cdot \left[\frac{1}{M}\sum_{i=1}^{M}\left(\theta^T x^{(i)} - y^{(i)}\right)x^{(i)}\right] \\ &= \frac{1}{M}\sum_{i=1}^{M}\left|x^{(i)}\right|^2 \geq 0\end{aligned}$$

There exist global minimum, our goal is to find it.

Multivariate Linear Regression

Feature Scaling:

$$x_j^{(i)} := \frac{x_j^{(i)} - \mu_j^x}{\sigma_j^x} \qquad \mu_j^x = \frac{1}{M}\sum_{i=1}^{M} x_j^{(i)} \qquad \sigma_j^x = \sqrt{\frac{\sum_{i=1}^{M}(x_j^{(i)} - \mu_j^x)^2}{M}}$$

Ensure all the features θ_j lie in the same range

$$J(\theta) = \frac{1}{2M}\sum_{i=1}^{M}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{2M}|X\Theta - Y|^2$$

$$= \frac{1}{2M}(\Theta^T X^T - Y^T)(X\Theta - Y)$$

$$= \frac{1}{2}(\Theta^T \frac{1}{M} X^T X\Theta - 2Y^T X\Theta + Y^T Y)$$

$$= \frac{1}{2}\Theta^T Q\Theta - b^T\Theta + c$$

$$Q = \frac{1}{M}X^T X \geq 0$$

$$b = \frac{1}{M}X^T Y$$

$$c = \frac{1}{2M}Y^T Y$$

Paraboloid

Multivariate Linear Regression

Structure of $$Q = \frac{1}{M} X^T X$$

$$\frac{1}{M} \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(M)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(M)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(1)} & x_N^{(2)} & \cdots & x_N^{(M)} \end{bmatrix}}_{\mathbb{R}^{(N+1)\times M}} \cdot \underbrace{\begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \cdots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \cdots & x_N^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(M)} & x_2^{(M)} & \cdots & x_N^{(M)} \end{bmatrix}}_{\mathbb{R}^{M\times(N+1)}}$$

$$= \underbrace{\begin{bmatrix} \frac{\sum_{i=1}^M 1}{M} & \frac{\sum_{i=1}^M x_1^{(i)}}{M} & \frac{\sum_{i=1}^M x_2^{(i)}}{M} & \cdots & \frac{\sum_{i=1}^M x_N^{(i)}}{M} \\ \frac{\sum_{i=1}^M x_1^{(i)}}{M} & \frac{\sum_{i=1}^M \left(x_1^{(i)}\right)^2}{M} & \frac{\sum_{i=1}^M x_1^{(i)} x_2^{(i)}}{M} & \cdots & \frac{\sum_{i=1}^M x_1^{(i)} x_N^{(i)}}{M} \\ \frac{\sum_{i=1}^M x_2^{(i)}}{M} & \frac{\sum_{i=1}^M x_2^{(i)} x_1^{(i)}}{M} & \frac{\sum_{i=1}^M \left(x_2^{(i)}\right)^2}{M} & \cdots & \frac{\sum_{i=1}^M x_2^{(i)} x_N^{(i)}}{M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^M x_N^{(i)}}{M} & \frac{\sum_{i=1}^M x_N^{(i)} x_1^{(i)}}{M} & \frac{\sum_{i=1}^M x_N^{(i)} x_2^{(i)}}{M} & \cdots & \frac{\sum_{i=1}^M \left(x_N^{(i)}\right)^2}{M} \end{bmatrix}}_{\mathbb{R}^{(N+1)\times(N+1)}} = \underbrace{\begin{bmatrix} \langle 1 \rangle & \langle x_1 \rangle & \langle x_2 \rangle & \cdots & \langle x_N \rangle \\ \langle x_1 \rangle & \langle (x_1)^2 \rangle & \langle x_1 x_2 \rangle & \cdots & \langle x_1 x_N \rangle \\ \langle x_2 \rangle & \langle x_2 x_1 \rangle & \langle (x_2)^2 \rangle & \cdots & \langle x_2 x_N \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle x_N \rangle & \langle x_N x_1 \rangle & \langle x_N x_2 \rangle & \cdots & \langle (x_N)^2 \rangle \end{bmatrix}}_{\mathbb{R}^{(N+1)\times(N+1)}}$$

Symmetric and positive semi-definite

Multivariate Linear Regression

Positive semi-definite $$v^TQv = v^TX^TXv = (Xv)^T(Xv) = u^Tu \geq 0$$

$$J(\theta) = \frac{1}{2M}\sum_{i=1}^{M}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{2M}|X\Theta - Y|^2$$

$$\nabla_\theta J(\theta) = \frac{1}{M}\left(\left(X^TX\right)\theta - \left(X^TY\right)\right)$$

$$Q = \frac{1}{M}X^TX \qquad b = X^TY$$

Normal Equation $$Q\theta \equiv \frac{1}{M}\left(X^TX\right)\theta = \frac{1}{M}X^TY$$

$$\theta = \left(X^TX\right)^{-1}X^TY = Q^{-1}b$$

$$O((N+1)^3)$$

Multivariate Linear Regression

Gradient descent method

Algorithm:

1. Set a precision ϵ, learning rate α, and set the initial guess θ_0
2. Let $\theta_{j+1} := \theta_j - \alpha \nabla_\theta J(\theta_j)$
3. Calculate $J(\theta_{j+1})$
    - If $|\theta_{j+1} - \theta_j| < \epsilon$, stop and return θ_{j+1}
    - Else, return to step 2

The algorithm is based on the Taylor expansion

$$J(\theta_{j+1}) = J(\theta_j - \alpha \nabla_\theta J(\theta_j)) = J(\theta_j) - \alpha(\nabla_\theta J(\theta_j))^2 + O(\alpha^2)$$