



Md. Ferdous · Follow

Feb 28, 2023



Maths behind Linear Prediction



I will start with basic loss calculation for linear prediction in this post. We know the loss equation is:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \theta_1 - x_i \theta_2)^2$$

If we convert input/features and output/labels as a matrix, then we can write:

$$J(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2$$

$$\begin{aligned}
J(\theta) &= (y - x\theta)^\top (y - x\theta) \\
&= y^\top y - y^\top x\theta - x^\top \theta^\top y + x^\top \theta^\top x\theta \\
&= y^\top y - y^\top x\theta - y^\top x\theta + x^\top \theta^\top x\theta \\
&= y^\top y - 2y^\top x\theta + x^\top \theta^\top x\theta
\end{aligned}$$

Now, we will calculate partial derivatives with respect to theta.

$$\begin{aligned}
\frac{\partial J(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} (y^\top y - 2y^\top x\theta + x^\top \theta^\top x\theta) \\
&= 0 - 2x^\top y + 2x^\top x\theta \\
&= -2x^\top y + 2x^\top x\theta
\end{aligned}$$

$$\text{Hence, } \frac{\partial A\theta}{\partial \theta} = A^\top \text{ and } \frac{\partial \theta^\top A\theta}{\partial \theta} = 2A^\top \theta$$

As we want to calculate θ accurately, hence

$$\frac{\partial \nabla(\theta)}{\partial \theta} = 0$$

$$\frac{\partial J(\theta)}{\partial \theta} = 0$$

$$- 2x^T y + 2x^T x \theta = 0$$

$$\Rightarrow 2x^T x \theta = 2x^T y$$

$$\Rightarrow \theta = \frac{x^T y}{x^T x}$$

$$\Rightarrow \theta = (x^T x)^{-1} x^T y$$

Now, we can calculate the correct theta using input/feature and output/label matrix. However, this simple linear prediction can lead to problems in poor data conditions. We need to add a regularizer to solve this issue; I will cover it in my next post.




Written by Md. Ferdous

1 Follower

Recommended from Medium

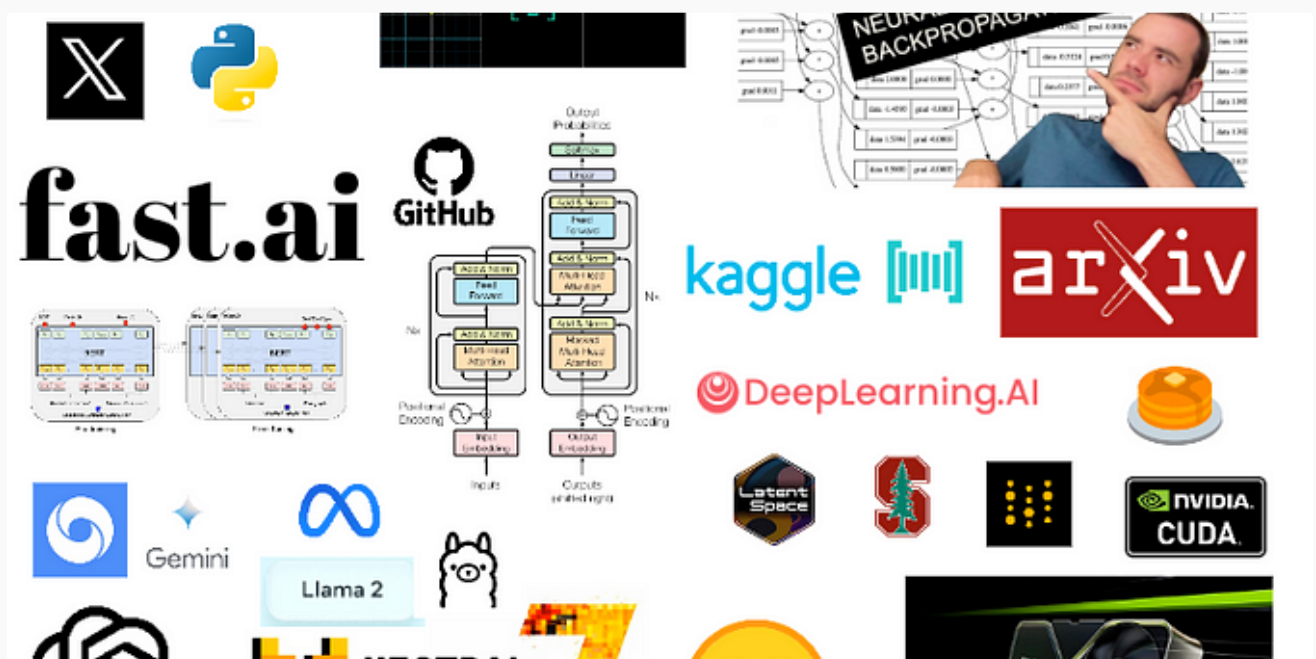


 Anandsullad

Exploring Linear Regression: A Practical Project

13 min read · Jan 11, 2024

 122 



Roadmap to Learn AI in 2024

A free curriculum for hackers and programmers to learn AI

11 min read · Mar 11, 2024



11.1K



121

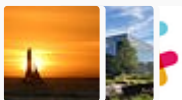


Lists



Staff Picks

637 stories · 962 saves



Stories to Help You Level-Up at Work

19 stories · 604 saves



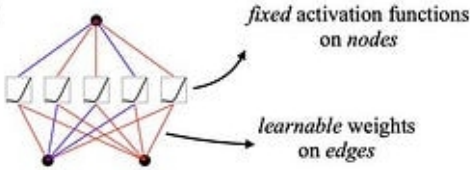
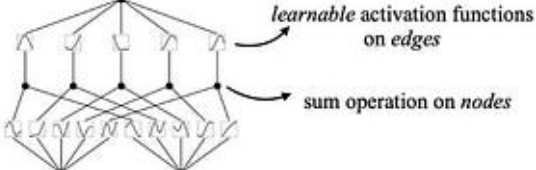
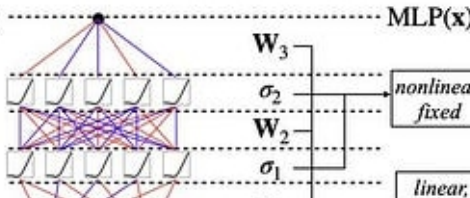
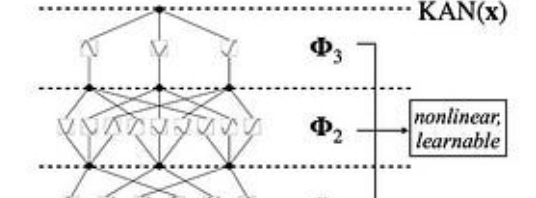
Self-Improvement 101


20 stories · 1785 saves



Productivity 101

20 stories · 1637 saves

Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

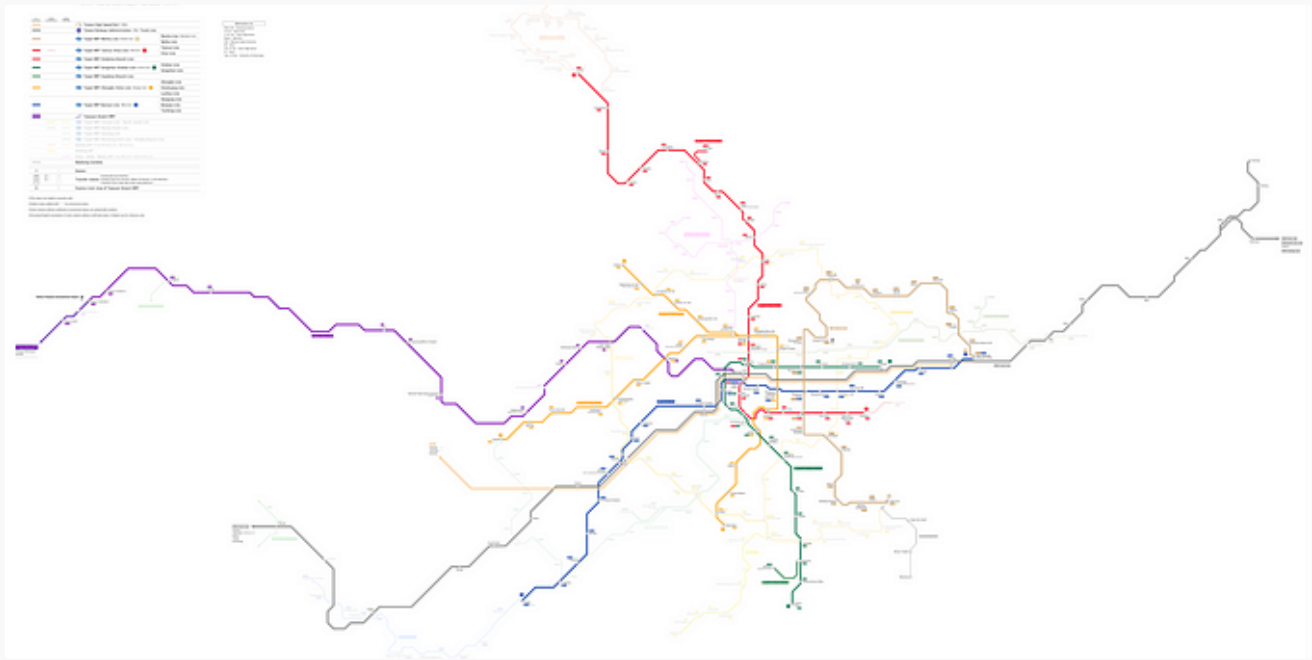
 Isaak Kamau

A Simplified Explanation Of The New Kolmogorov-Arnold Network (KAN) from MIT

Exploring the Next Frontier in AI: The Kolmogorov-Arnold Network (KAN)

2 min read · May 1, 2024

 1.8K  4



 Beth Ou Yang in Towards Data Science

PCA & K-Means for Traffic Data in Python



Reduce dimensionality and cluster Taipei MRT stations based on hourly traffic

8 min read · 3 days ago

 186  1





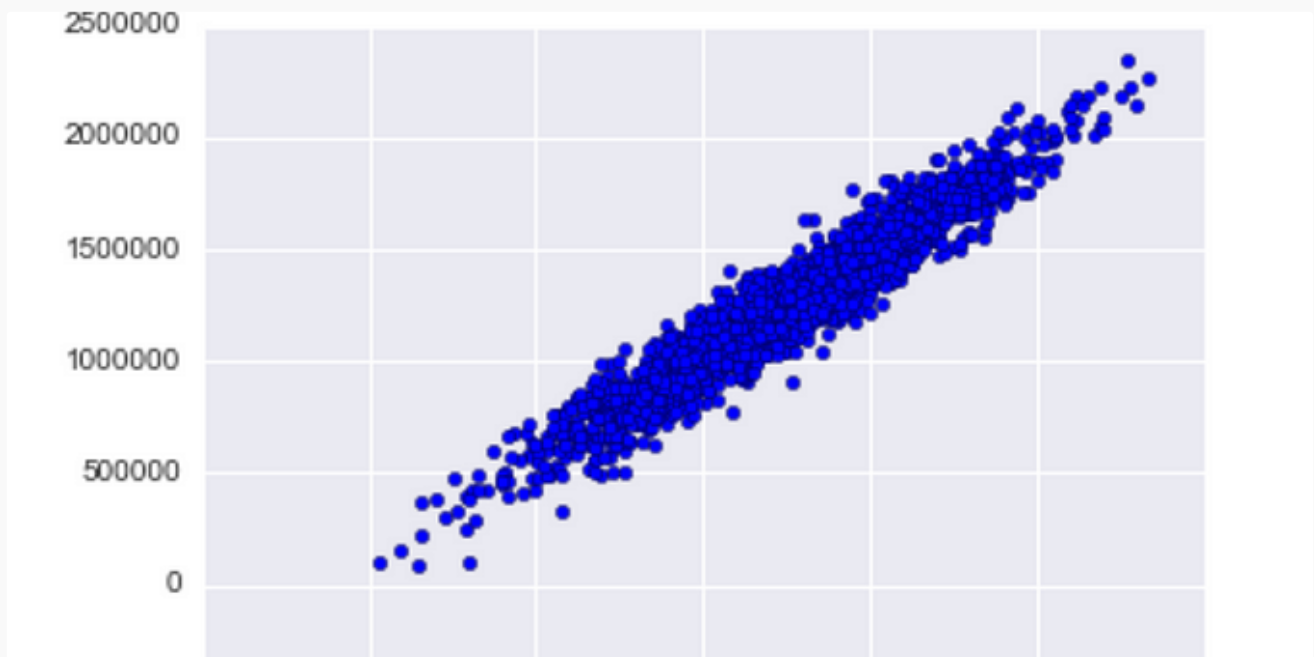
 Vishal Rajput  in AIGuys


Can Kolmogorov–Arnold Networks (KAN) beat MLPs?

A completely new way to think about Neural Networks based on Spline based function approximation.

🌟 · 9 min read · 5 days ago

 313  2



 Simran Kaushik

House Price Prediction: A Simple Guide with Scikit-Learn and Linear Regression

Navigate the realm of predictive analytics with simplicity

7 min read · Nov 14, 2023



89



1



See more recommendations