Artificial Intelligence Club Week 2 Slides

October 8, 2025

TECH TALK

OSU AI CLUB

JOIN US IN-PERSON FOR AIC TECH TALK

Location: *KELLEY 1001 (live-stream on Zoom)*
Date: *13 October 2025, 6:30pm*

David Zier

NVIDIA	2025-PRESENT
director of deep learning system software	
NVIDIA	2021-25
senior manager	
NVIDIA	2018-2021
system software manager	
OREGON STATE	2002-09
graduate teaching assistant	

LEARN MORE

@osu_aiclub

Oregon State University AI Club

Linear Algebra … Why do we need it?

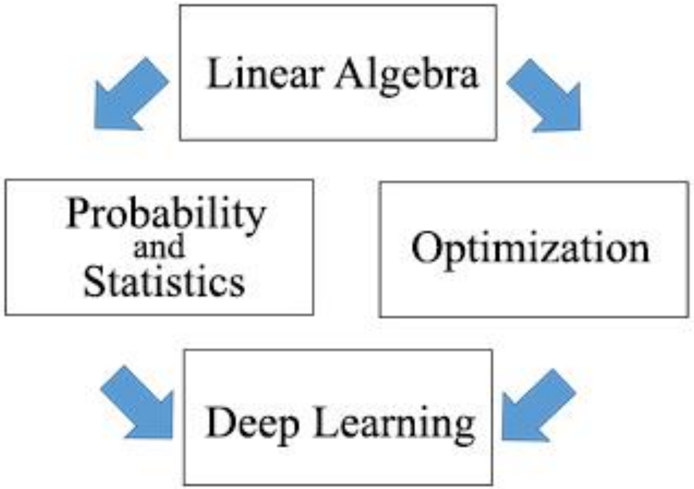
Turns words and images into numbers and organizes them!

Our models can now understand the input!

Variables ->

<- Events

	Var 1	Var 2	…	Var b
1	a_11	a_12	…	a_1b
2	a_21	a_22	…	a_2b
…	…		…	…
n	a_n1	a_n2		a_nb



Linear Algebra ... The Basics

Scalar	Vector	Matrix	Tensor
1	$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$	$\begin{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} & \begin{bmatrix} 3 & 2 \end{bmatrix} \\ \begin{bmatrix} 1 & 7 \end{bmatrix} & \begin{bmatrix} 5 & 4 \end{bmatrix} \end{bmatrix}$

$$
\begin{array}{c c}
 & \begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array} &
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
a_{31} & a_{32} & \cdots & a_{3n} \\
\vdots & \vdots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\end{array}
$$

Matrix rules

scalar multiplication	$n \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = \begin{bmatrix} na & nb & nc \\ nd & ne & nf \end{bmatrix}$
matrix addition	$\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} + \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} a+g & b+h \\ c+i & d+j \\ e+k & f+l \end{bmatrix}$
matrix multiplication	$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} ag+bi+ck & ah+bj+cl \\ dg+ei+fk & dh+ej+fl \end{bmatrix}$

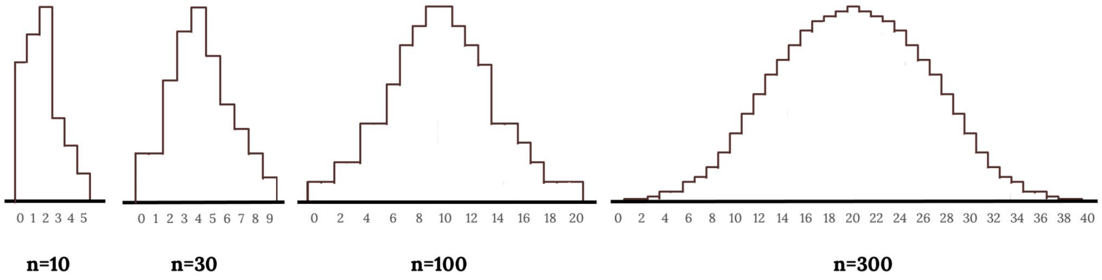
Statistics … Why do we need it

Almost all AI/ML agents/models are predictive!

- They select the output that is most likely to happen, based on the data it was trained on, what it was training to value, and the type of model it is

AI upgrades what humans are capable of

- Humans have always came up with conclusions and predictions based on the data we were given, but some processes are too complex for a human to be able to predict or explain



Statistics … Basics

- **Expectation aka the Mean:**

  The expectation is the every possible output multiplied by its probability, then summed

- **Variance:**

  The variance is the average squared distance between each data point and its mean

- **Statistical Distributions:**

  Different models to describe the possible values of a variable and the frequency or probability of each value occurring

  - **Discrete and Continuous**

    Discrete values are finite, like which side of a die is rolled

    Continuous results are infinite, like choosing a real number between 1 and 100

But how do both of these relate back to predicting outcomes using data?

Bayesian vs Frequentist

Lebron is about to shoot a free throw, what are the chances he makes it?

Frequentist: So far in the game, he shot 5/10 free throws. This means his free throw percentage is 50%, so **50% chance** he makes it!

Bayesian: However, Lebron has also shot 200 free throws so far this season, scoring 170 of them. This information can be used to come up with a better prediction. We need to include this information/data, also called a **prior**.

We use a Beta function to do that, where:

Beta(prior success + data success, prior misses + data misses) -> Beta(170 + 5, 30 + 5) = Beta(175, 35) = 175 / 175 + 35=0.833

Thus, the predicted probability (posterior mean) is 83%, giving us an 83% chance he makes it!

Bayesian vs Frequentist

The **Frequentist** standpoint emphasizes how Lebrons free throw percentage changes with each **game**, thus we try to predict the chances of his next shot going in with data from the **current game.**

The **Bayesian** standpoint emphasizes that historical data can also guide our prediction. It becomes even more powerful when we include **scaling**, a way to add more **human emphasis** on how much influence we think the historical data should have. Say Lebron is sick this game, so the prior is less reliable:

Beta(scaled prior success + data success, scaled prior misses + data misses) $\rightarrow$
Beta((170 × 0.5) + 5, (30 × 0.5) + 5) = Beta(90, 20) =0.818

This gives us an 81.8% chance he makes it, slightly lower than using the full historical data, since we scaled the prior's influence to 50%. The scaling acts like a "confidence dial" on how much we trust the past versus what we've just observed.

We do this all the time naturally!

Our brains automatically consider historical data, when it comes to making future predictions!

But these ideas also need to be formalized and proven mathematically

Suppose we have data $\mathcal{D} = \{x^{(i)}\}_{i=1}^{N}$

Maximum Likelihood Estimate (MLE)

$$\boldsymbol{\theta}^{\text{MLE}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta})$$

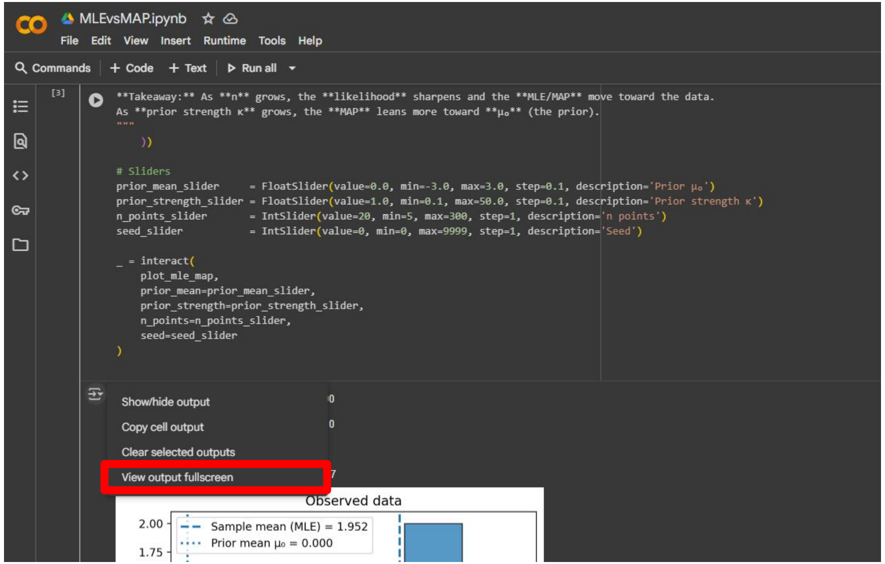
Maximum *a posteriori* (MAP) estimate

$$\boldsymbol{\theta}^{\text{MAP}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta})\underbrace{p(\boldsymbol{\theta})}_{\text{Prior}}$$

Maximum Likelihood Estimation VS Maximum a Posteriori

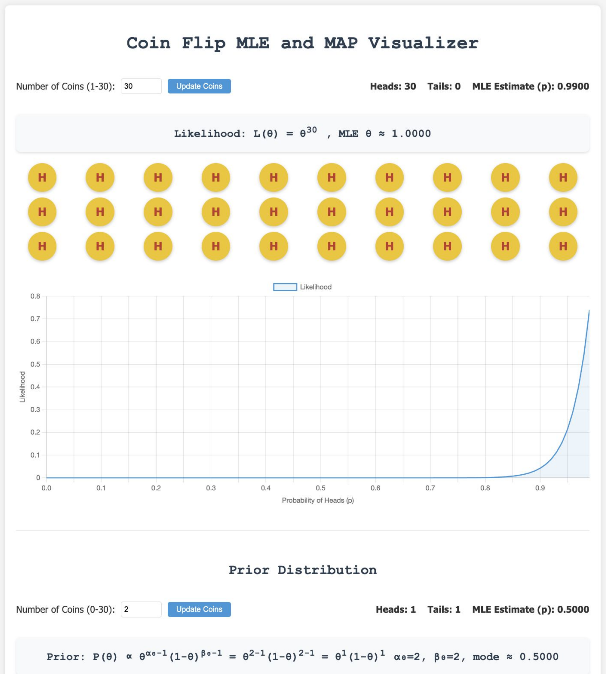
https://colab.research.google.com/drive/16TUgd1C4QxBQFYOi0fq1dTOvWbgcNeLT?usp=sharing

(Make sure to hit view output fullscreen)



Discrete MLE and MAP Example with Coin Flips

https://mahmoudfakhry.github.io/MLE_and_MAP/



Similarities, Differences, & Philosophies

- **Similarities:** Both use the **likelihood**; both give point **estimates**; often **coincide as n grows** (prior influence fades).

- **Differences: MLE ignores priors** (can be high-variance, boundary estimates); **MAP uses a prior** (adds bias, reduces variance).

- **Frequentist (MLE):** Parameters are fixed based on current information.
  Uses random sampling to generate parameters for prediction.
  a. EX: Asking 10,000 people who they will vote for and using that to predict election

- **Bayesian (MAP):** Parameters are random and depend on the size of the prior, as well as the size of the data! **prior + data → posterior**
  a. EX: Asking those 10,000 people, but also considering historical voting

Resources

- ## Coding/Python:

  Easier: https://www.codecademy.com

  More complex: https://www.youtube.com/watch?v=kqtD5dpn9C8&ab_channel=ProgrammingwithMosh

  and many many others

- ## Linear Algebra:

  Easier: https://youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab&si=qQnVeyJd58BkU4AV

  More complex: https://youtu.be/N1Pvj4CZT1M?si=PbvkwWiJlulsgfLD

  In machine learning: https://www.visual-design.net/post/linear-algebra-for-machine-learning

- ## Statistics:

  Easier: https://www.youtube.com/watch?v=NIqeFYUhSzU

  More complex: https://www.youtube.com/watch?v=WB8eYZSZyaE

Connections + Pizza + Talk to Officers

FEEDBACK??